Acari genomics: from phylogenetics to agricultural applications

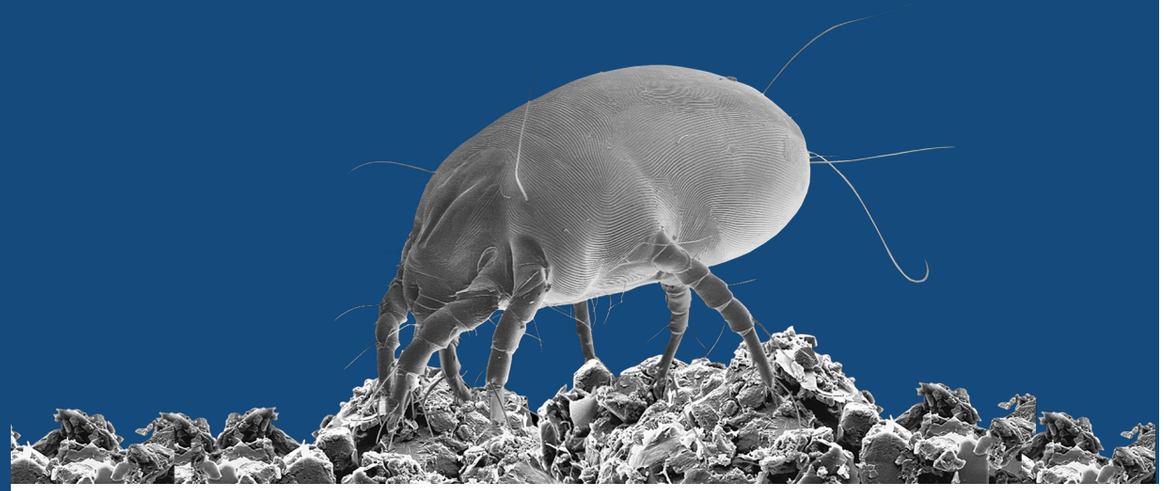
Wannes Dermauw

2013

ISBN 978-9-0598960-8-6

Acari genomics: from phylogenetics to agricultural applications

Ir. Wannes Dermauw



UNIVERSITEIT GENT

FACULTEIT BIO-INGENIEURSWETENSCHAPPEN

Man is certainly crazy. He could not make a mite, and he makes gods by the dozen.
(Michel de Montaigne 1533-1592)

It only takes five minutes…
(Thomas Van Leeuwen)

Promoters: **Prof. dr. ir. Luc Tirry**
Department of Crop Protection
Faculty of Bioscience Engineering
Ghent University

Dr. ir. Thomas Van Leeuwen
Department of Crop Protection
Faculty of Bioscience Engineering
Ghent University

Dean: **Prof. dr. ir. Guido Van Huylenbroeck**

Rector: **Prof. dr. Paul Van Cauwenberge**

ACARI GENOMICS: FROM PHYLOGENETICS TO AGRICULTURAL APPLICATIONS

ir. Wannes Dermauw

Thesis submitted in fulfillment of the requirements for the degree of
Doctor (PhD) in Applied Biological Sciences

Dutch translation of the title of this PhD:
Acari genomica: van fylogenetische analyses tot toepassingen in de landbouw

To refer to this thesis:

Dermauw W (2013). Acari genomics: from phylogenetics to agricultural applications. PhD thesis. Ghent University, Ghent, Belgium.

ISBN-number: 978-90-5989-608-6

Front cover: top: ESEM photograph of the front part of the two-spotted spider mite, *Tetranychus urticae* (© Wannes Dermauw); bottom: ESEM photograph of the European house dust mite *Dermatophagoides pteronyssinus* (© David McCarthy); middle: DNA-helix structure.
Back cover: top left corner: SEM photograph of the predatory mite *Phytoseiulus persimilis* (© David Evans Walter)

The promoters, The author

Prof. Dr. ir. Luc Tirry Dr. ir. Thomas Van Leeuwen ir. Wannes Dermauw

Members of the examination committee

Prof. dr. ir. Luc Tirry
Laboratory of Agrozoology
Department of Crop Protection
Ghent University

Dr. ir. Thomas Van Leeuwen
Laboratory of Agrozoology
Department of Crop Protection
Ghent University

Prof. dr. ir. Kris Verheyen
Forest & Nature Lab
Department of Forest and Water Management
Ghent University

Prof. dr. Lieve Gheysen
Laboratory of Applied Molecular Genetics
Department of Molecular Biotechnology
Ghent University

Prof. dr. ir. Tim De Meyer
Department of Mathematical Modelling, Statistics and Bioinformatics
Ghent University

Prof. dr. John Vontas
Laboratory of Molecular Entomology
Department of Biology
University of Crete, Greece

Dr. René Feyereisen
INRA - CNRS- Université Nice Sophia Antipolis
Centre de Recherches de Sophia Antipolis, France

VOOR/DANKWOORD

“Je kan toch niet de rest van je leven parkeerlijnen blijven schilderen” zei mijn moeder in de zomer van 2006. Ze had ergens wel een punt. Na wat opzoekwerk viel mijn oog op een vacature bij het labo Agrozoölogie van professor Luc Tirry. De leuke herinneringen van mijn thesis in dit labo in gedachten, solliciteerde ik er voor de functie van assistent dierkunde. Ik kreeg de job en begon aan een nieuw avontuur.

Ondertussen naderen we de zomer van 2013 en een heuse doctoraatsthesis ligt voor jullie neus. Geen thesis echter zonder resultaten. Elke doctoraatsstudent weet dat het behalen van dat laatste niet altijd evident is, maar een goede begeleiding kan wonderen doen. Luc, en Thomas, zonder jullie was dit doctoraat er nooit gekomen. Luc, bedankt dat je me de kans gaf om een doctoraat te beginnen en steeds bereid was om mij met raad en daad, op eender welk vlak, bij te staan. Was het niet wat betreft de determinatie van een kever van het genus *Cetonia* (of was het *Netocia* ?*)* in een pak diepvries broccoli dan wel met hoe je een goede presentatie moet opbouwen. Thomas, waar zal ik beginnen? Je niet aflatende enthousiasme voor wetenschap werkte heel stimulerend. Weinig wetenschappelijke (en levensbeschouwelijke) vragen bleven bij jou onbeantwoord, maar dan wel liefst na een goede nachtrust, een Gauloise en, niet te vergeten, een blik cola ;-). Via jouw contacten kon ik ook 3 maanden ervaring opdoen in een labo in Kreta, een unieke belevenis die ik mij nog lang zal herinneren. Voor dit alles en nog veel meer, oprecht bedankt. Je zal het ongetwijfeld (rookvrij?) uitstekend doen in Amsterdam!

In Kreta heb ik drie onvergetelijke maanden beleefd, zowel in het labo als daarbuiten. John, Evangelia en Aris, bedankt voor jullie fantastische gasthvrijheid. Ας ελπίσουμε ότι, θα συναντηθούν σύντομα στην Κρήτη. Maar ook in het labo aan de Coupure was het fijn vertoeven. In het bureau van de assistenten, eerst samen met Bjorn en Steven, daarna met Jochem en Astrid. Mijn oprechte excuses als mijn koffie soms iets te lang stond te beschimmelen of als ik de deur weeral vergeten te sluiten was! Jochem en Astrid, het verschuiven van bureaus kan nu eindelijk beginnen ;-). Onder het motto “afzien in groep is leuker dan alleen”, ben ik ook veel dank verschuldigd aan mijn collega’s van de mijtenclub. Pieter, Jahangir, Peter, Nicky, Astrid, Wim, Arash en Sabina, zonder jullie was het mijtenonderzoek maar een eenzame aangelegenheid geweest! Speciaal voor Jahangir: اب رد دوخ هدافلا قوف یزاون نامم نانوی رد نم هوق تخاس یراب امش زا رکشت.نانوی رد نم هوق تخاس یراب امش زا رکشت اریان اسب.اب هبترین آرزوها براي امش و خانداوتان.

Ook personen buiten de "*Acarology crew"* hebben mij goed geholpen. Bedankt Bartel Vanholme om mij in mijn eerste doctoraatsjaren wegwijs te maken in de wereld van de moleculaire biotechnologie en de bioinformatica (en om je bench met ons labo te delen ;-)). Joachim Jacob, je Southern blot (en Corel Draw) skills zijn mij heel goed van pas gekomen tijdens mijn doctoraat! Thijs, merci om tot vervelens toe competente cellen te gaan halen en voor je goede vinyl- en metseltips. Nick en Maarten, zonder jullie geen leuke labo-uistapjes! Rik, Leen en Didier, zonder jullie technische en administratieve hulp zou mijn doctoraat niet zo vlot verlopen zijn, bedankt!

Aan de leden van mijn PhD-commissie: John, thank you very much for the excellent feedback and cooperation. René, from now on the Fifth of Beethoven will always sound different to me. Your help was indispensable for the success of my PhD. Bedankt ook aan Prof. Verheyen, Prof. Gheysen en Prof. De Meyer voor de positieve commentaar en het nauwkeurig nalezen van mijn doctoraat.

In die zeven jaar was er natuurlijk ook een leven buiten de wetenschap. Wat zou een mens doen zonder zijn vrienden? Jullie zorgden voor de ideale afleiding tijdens mijn doctoraat! Ondertussen zijn de meeste van jullie vanuit Gent uitgezwermd naar Brussel, Kortrijk, Oostakker of zelfs London. Na mijn doctoraat moeten we er dringend werk van maken om elkaar eens meer te zien! Beloofd!

Ook mijn ouders waren essentieel in het slagen van mijn doctoraat. Ze boden mij de kans om te studeren en steunden mij in alles wat ik deed. Vake, geef toe, mijtendokter klinkt ook niet slecht ;-). Daarnaast mag ik natuurlijk mijn twee zusjes, Liselotte en Annelore, niet vergeten. Jullie zijn er in geslaagd om allebei een goede man te pakken te krijgen, en zorgden dat ik twee neefjes en twee nichtjes erbij kreeg. Bruna en Oskar, ik ben trots om jullie peter te zijn. Ondertussen is er tijdens mijn doctoraat ook een hele schoonfamilie bijgekomen en ook op dat vlak heb ik het getroffen. Patrik en Ann, jullie gastvrijheid zowel in Kortrijk als in Lissoir zijn ongeëvenaard!

Lieve Nathalie, natuurlijk wordt de laatste paragraaf van dit dankwoord aan jou opgedragen. Eindelijk trekken we er samen weer op uit voor een groot reisavontuur. Reeds zes jaar samen, reeds zes jaar een feest met jou. Jouw optimisme en niet aflatende energie hielpen mij door die laatste moeilijke doctoraatsmaanden. Bedankt voor alles!

Wannes, mei 2013

Table of contents

List of Abbreviations

12S-rRNA	small (*12S*) *rRNA* subunit
16S-rRNA	large (*16S*) *rRNA* subunit
A	*a*denine
AA	*a*mino acid
ABC	*A*TP-*b*inding *c*assette
ATP	*a*denosine *t*ri*p*hosphate
atp6 and *8*	*ATP*ase subunit 6 and 8
BI	*b*ayesian *i*nference
BLAST	*b*asic *l*ocal *a*lignment *s*earch *t*ool
bp	*b*ase *p*airs
C	*c*ytosine
CCE	*c*arboxyl/*c*holine *e*sterase
cDNA	*c*omplement *DNA*
CFTR	*c*ystic *f*ibrosis *t*ransmembrane conductance *r*egulator
CI	*c*onfidence *i*nterval
cox1-3	*c*ytochrome *ox*idase subunits *I-III*
cy3/5	*cy*anine *3/5*
CYP	*cy*tochrome *P*450 mono-oxygenase
cysLGIC	*cys*-loop *l*igand-*g*ated *i*on *c*hannel
cytB	*cyt*ochrome *b*
D-arm	*d*ihydrouridine-arm of a tRNA secondary structure
Da	*Da*lton
DD	*d*egree of *d*ominance
df	*d*egrees of *f*reedom
DIG	*dig*oxygenin
DMSO	*dim*ethyl*s*ulf*o*xide
DNA	*d*eoxyribo*n*ucleic *a*cid
dNTP	*d*eoxyribo*n*ucleotide *t*ri*p*hosphate
dsRNA	*d*ouble *s*tranded *RNA*
EC	*e*mulsifiable *c*oncentrate
EDTA	*e*thylene*d*iamine*t*etraacetic *a*cid
EF1-α	*e*longation *f*actor 1-α
EST	*e*xpressed *s*equence *t*ag
EtdBr	*et*hi*d*ium *b*romide
FC	*f*old *c*hange
FDR	*f*alse *d*iscovery *r*ate
FT	*f*ull *t*ransporter
G	*g*uanine
GABA	*γ-a*mino *b*utyric *a*cid
GE	*g*ene *e*xpression
GluCl	*glu*tamate-gated *c*h*l*oride channels
GST	*g*lutathione *S-t*ransferase
HisCl	*his*tamine-gated *c*h*l*oride channels
HT	*h*alf *t*ransporter
ID-RCD	*i*ntradiol *r*ing-*c*leavage *d*ioxygenase
J-strand	ma*j*ority strand
kb	*k*ilo*b*ase
L:D	*l*ight:*d*ark
LC_{50}	*l*ethal *c*oncentration that kills 50% of a population
lnr	*l*arge *n*on-coding *r*egion/control region
LS-VL	*l*aboratory *s*usceptible strain *V*an *L*eeuwen
MAR-AB	*mar*athonas *ab*amectin resistant strain
MDR	*m*ulti*d*rug *r*esistance protein
METI	*m*itochondrial *e*lectron *t*ransport *i*nhibitor
MFS	*m*ajor *f*acilitator *s*uperfamily
ML	*m*aximum *l*ikelihood
MP	*m*aximum *p*arsimony
mRNA	*m*essenger *RNA*
MRP	*m*ultidrug *r*esistance associated *p*rotein
MR-VP	*M*ETI-*r*esistant strain *V*an *P*ottelberghe
mt	*mit*ochondrial
N-strand	*m*inority *s*trand
nAChR	*n*icotinic *a*cetyl*ch*oline *r*eceptors
nad1-6	*NAD*H dehydrogenase subunits *1–6*
nad4L	*NAD*H dehydrogenase subunit *4L*
NADH	*n*icotinamide *a*denine *d*inucleotide
NBD	*n*ucleotide *b*inding *d*omain
NCBI	*N*ational *C*enter for *B*iotechnology *I*nformation
ORF	*o*pen *r*eading *f*rame
P-gp	*P-g*lyco*p*rotein
P450	cytochrome *P450* mono-oxygenase
PCG	*p*rotein *c*oding *g*ene
PCR	*p*olymerase *c*hain *r*eaction
pHCl	*pH*-sensitive *c*h*l*oride channel
ppm	*p*arts *p*er *m*illion, mg active ingredient per litre
qPCR	*q*uantitative realtime *PCR*
R.H.	*r*elative *h*umidity
RACE	*r*apid *a*mplification of *c*DNA *e*nds
rdl	*R*esistance to *d*ie*l*drin
RISC	*R*NA-*i*nduced *s*ilencing *c*omplex

RNA	*r*ibo*n*ucleic *a*cid
RNAi	*RNA i*nterference
RPKM	*r*eads *p*er *k*ilobase per *m*illion mapped reads
rpm	*r*ounds *p*er *m*inute
RR	*r*esistance *r*atio
rRNA	*r*ibosomal *RNA*
RSCU	*r*elative *s*ynonymous *c*odon *u*sage
SC	*s*uspension *c*oncentrate
SDS	*s*odium *d*odecyl *s*ulfate
SE	*s*tandard *e*rror
siRNA	*s*hort *i*nterfering *RNA*
SUR	*s*ulfonyl*u*rea *r*eceptor
T	*t*hymine
T-arm	*T*Ψ*C-arm* of a tRNA secondary structure
TAE	*T*RIS-*a*cetate *E*DTA
TBE	*T*RIS-*b*orate *E*DTA
TM1-4	*t*rans*m*embrane region 1 to 4
TRIS	*tris*(hydroxymethyl) - aminomethane
trnX	*tr*a*n*sfer RNA (where *X* is replaced by a one letter amino acid code for the corresponding amino acid, with L_1=CUN; L_2=UUR; S_1= AGN; S_2=UCN)
UTR	*u*n*t*ranslated *r*egion
WP	*w*ettable *p*owder

Scope

Acari, mites and ticks, are a very diverse and species rich group of animals. Compared to insects however, their genomes are underexplored. This dissertation aims at delivering a contribution to the emerging field of Acari genomics.

In **Chapter II** and **III**, a long PCR strategy was employed to obtain the mitochondrial genomes of two economically important mites, the European house dust mite *Dermatophagoides pteronyssinus* and the predatory mite *Phytoseiulus persimilis*. Both mite mitochondrial genomes were profoundly analyzed and a restriction digest of rolling circle amplified mitochondrial DNA was performed to confirm their length. Furthermore, the usefulness of mitochondrial genome orders for deducting Acari relationships was evaluated.
In 2011 the complete nuclear genome of the spider mite *Tetranychus urticae* was sequenced. This mite species is one of the most polyphagous herbivores known to man and considered as the "pesticide resistance champion" among arthropods. The availability of its genome provided an excellent opportunity to study the molecular mechanisms behind these phenomena. A whole-genome gene expression microarray was developed and used to follow changes in *T. urticae* gene expression after transfer to a new challenging host, tomato, and in pesticide-susceptible versus pesticide-resistant strains. Striking similarities were found between the gene families and transcriptional responses associated with plant host adaptation and pesticide resistance, thereby shedding light on the nature of pesticide resistance (**Chapter IV**).

In **Chapter V** the ABC protein family, known to be involved in detoxification of xenobiotics, was thoroughly characterized in *T. urticae*. Phylogenetic relationships were inferred between *T. urticae* ABC proteins and their metazoan counterparts and the expression profile of ABC genes either upon host transfer or in pesticide resistant strains was determined. Another important gene family in resistance research are the cys-loop ligand gated ion channels (cysLGICs). Members of this family are known as possible target-sites of pesticides and were hence catalogued for *T. urticae* (**Chapter VI**). Phylogenetic relationships between *T. urticae* cysLGICs and their insect counterparts were inferred and toxicological implications associated with specific features of these receptors in *T. urticae* compared to insects were discussed. In addition, the putative role of these channels in abamectin resistance was further

investigated and revealed that an accumulation of mutations in different glutamate gated chloride channel genes is associated with high levels of abamectin resistance.

Finally, in **Chapter VII**, general conclusions and future perspectives based on the results obtained in this study are provided.

Chapter I

Introduction to Acari Genomics

1 Acari

Acari - mites and ticks - belong to the subphylum of the Chelicerata, the second largest group of terrestrial animals and one of the most diverse groups of animals in the world (Fig. I.1). They diverged 400 million years ago and comprise almost 55 000 described species, but the actual number of species is thought to be twentyfold higher (Table I.1). Some feed on plants, bacteria or fungi while others have developed obligate associations with vertebrate and invertebrate animals. They are capable of colonizing a range of habitats, exceeding those known for insects: from the tundra to hot deserts, from oceanic trenches to our own eyebrows ((Walter and Proctor 1999; Krantz and Walter 2009; Dunlop 2010), Table I.1). Moreover, some mite species exhibit properties one could never have imagined: from superfast backward jumping to lifting 1,200 times their own weight, from being the source of poison frog's deadly defenses to creating a unique aroma in cheese (Hase 1929; Wauthy *et al.* 1998; Heethoff and Koerner 2007; Raspotnig *et al.* 2011). Who knows which extraordinary features are hidden in all the species remaining to be discovered?

1.1 Classification

Acari belong to the Arthropoda. This group of animals comprises both terrestrial and aquatic invertebrates that are characterized by the presence of jointed appendages and a chitinous exoskeleton. Within this phylum, the subphylum of the Chelicerata can be discerned from other subphyla by the absence of a separate head, antennae or wings and the presence of a pair of pincer- or stylet like mouthparts, called chelicerae, and, in most cases, 4 pairs of legs. The largest chelicerate class is the Arachnida, of which the subclass of the Acari (mites and ticks) and the orders of the Araneae (spiders) and Scorpiones (scorpions) are the most well known. Acari can be distinguished from other Arachnida due to their small size, in many cases less than 1 mm, and the reduction of their body segmentation to 1 complex. This complex is artificially divided in the *gnathosoma* or *capitulum*, carrying the mouthparts, and the *idiosoma*. The Acari are divided into two superorders: the Acariformes or Actinotrichida and the Parasitiformes or Anactinotrichida. The first group is characterized by the presence of actinopilin in their setae, causing birefringence under polarized light, and comprises two species rich orders, the Sarcoptiformes (containing a.o. stored product mites, mange mites and house dust mites) and the Trombidiformes (with most plant feeding mites). The Parasitiformes group mainly consists of Ixodida (ticks) and Mesostigmata (e.g. the predatory

phytoseiid mites) (Table I.1, Fig. I.2). The Acari are classified in 124 superfamilies, 540 families and 5500 genera (Walter and Proctor 1999; Krantz and Walter 2009).

Figure I.1 – Multipanel displaying the diversity of Acari.
A: a spider mite, *Tetranychus urticae*, B: the house dust mite, *Dermatophagoides pteronyssinus*, C: a box mite, *Rhysotritia duplicata*, D: an engorged hard tick, *Ixodes ricinus*, E: the honeybee mite, *Varroa destructor*, F: leaf galls caused by *Eriophyes tiliae*; G: a predatory mite, *Phytoseiulus persimilis*; H: a nematode like mite *Gordialycus tuzetae*; I: a red velvet mite (Trombidiidae) (copyright A, B, E: Gilles San Martin, C: Bartel Vanholme, D: Richard Bartz; F: Roger Griffith; G: David Evans Walter; H: Haupt and Coinau 1999; I: Thomas Shahan)

It is noteworthy that the monophyletic origin of the Acari has been long-debated and is to date the subject of discussion (reviewed by Dunlop and Alberti 2008). Van der Hammen (1989) suggested a diphyletic origin of Acari and placed the Parasitiformes closest to Ricinulei (hooded tickspiders) while Acariformes were related to the Palpigradi (microwhip scorpions) (Fig I.2). Lindquist (1984, 2009), on the other hand, defended, a monophyletic origin of Acari based on the fact that Acariformes and Parasitiformes share a variety of presumable apomorphies (a derived trait, that is unique to a given terminal group) and because there is, according to Lindquist, no convincing evidence that these lineages are more closely related to any other group of Arachnida than they are to each other. Overall, the hypothesis of Lindquist has been the most favoured one (Dunlop and Alberti 2008).

However, this favoured hypothesis has recently been rejected by several molecular studies, with Acariformes being grouped with either Solifugae (camel spiders) or Palpigradi, and Parasitiformes being related to either Pseudoscorpiones or Opiliones (harvestmen) (Fig. I.2) (Dabert *et al.* 2010; Pepato *et al.* 2010; Regier *et al.* 2010; Rota-Stabelli *et al.* 2013). Nevertheless, the phylogenetic analyses in these studies contained either a low number of Acari taxa or were based on only few gene sequences, so a more thorough phylogenomic approach (see 3.1 Phylogenetics) should be performed before a conclusive answer can be found for this enigmatic phylogenetic issue.

Table I.1 – Total number of described Acari species and their maximum estimated numbers (Walter and Proctor 1999; Zhang 2011)

group	number of described species	maximum estimated number
Acari	54,617	1,132,090
Parasitiformes	12,359	202,190
Opilioacarida	35	170
Holothyrida	27	320
Ixodida	873	1,200
Mesostigmata	11,424	200,500
Acariformes	42,258	929,900
Trombidiformes	25,845	637,500
Sarcoptiformes	16,413	292,400

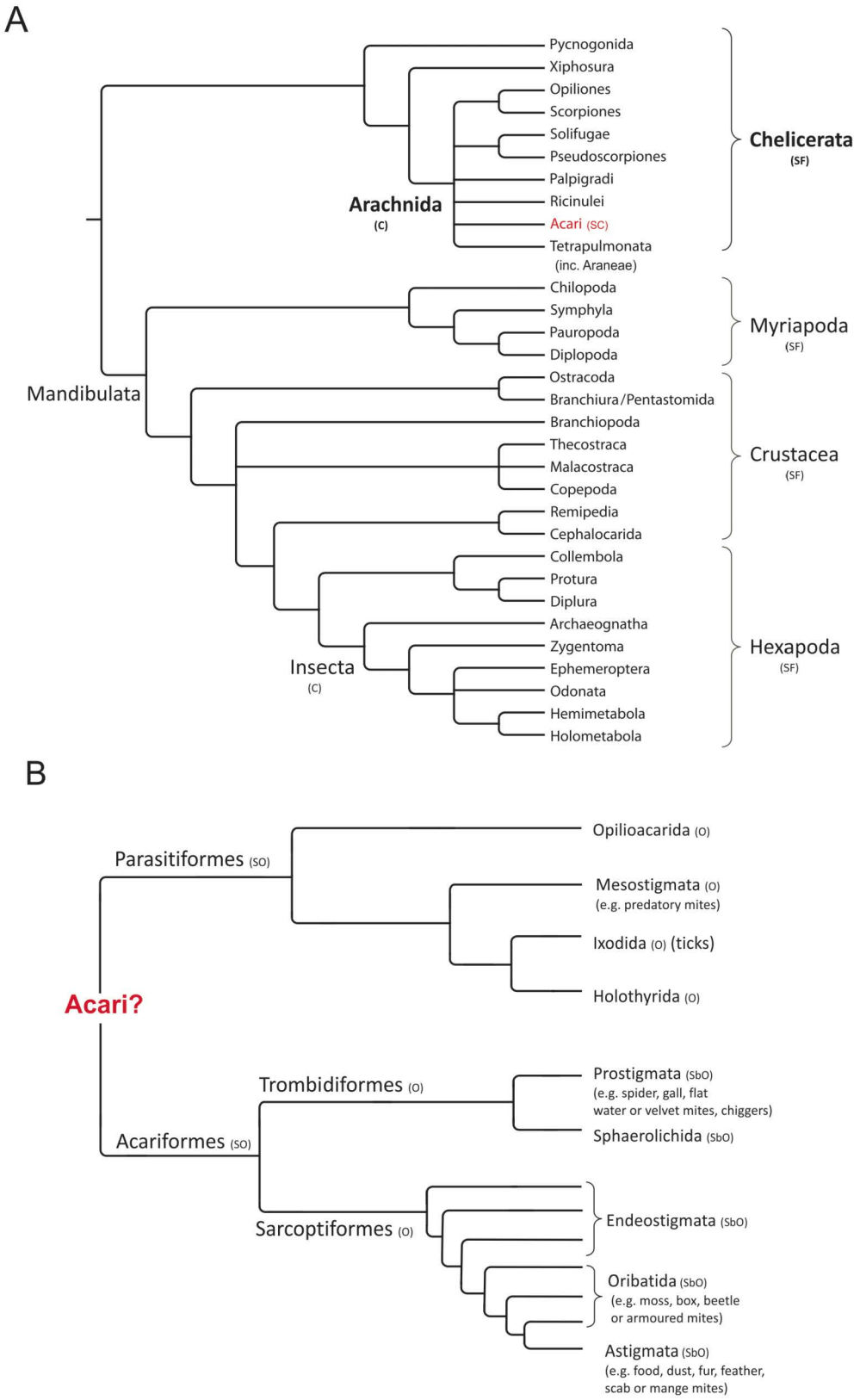


Figure I.2 – Phylogeny of the Acari

A) Phylogeny of the Arthropoda phylum according to the Mandibulata hypothesis; the position of the Acari is indicated in red (drafted from Giribet and Edgecombe 2012); B) Phylogeny of the Acari; the monophyly of the Acari is still under debate (drafted from Dunlop and Alberti 2008 and Krantz and Walter 2009). Abbreviations of taxonomic rank: SF: subphylum; C: class; SC: subclass; SO: superorder; O: order; SbO: suborder.

1.2 Acari of agricultural and medical importance: a brief survey

Compared to insects, mites have often been treated like cinderellas. The aim of the following survey is to give an idea of the overall importance of this group of tiny arthropods.

Acari species of agricultural importance exhibit either phytophagous (plant feeding), predatory of parasitic habits. The majority of agriculturally important phytophagous mites belong to the superfamilies of the Tetranychoidea and Eriophyoidea, both belonging to the Trombidiformes (Krantz and Lindquist 1979; Lindquist *et al.* 1996; Van Leeuwen *et al.* 2010a). Within the Tetranychoidea, the Tetranychidae or spider mite family contains the economically most important mite crop pests. The annual cost of their chemical control is estimated at 0.5-1 billion dollars worldwide (McDougall 2008). These mites use their stylet like chelicerae to puncture the leaf mesophyll-cells and suck up the cell content, resulting in chlorotic spots on the leaf (Jeppson *et al.* 1975). The best known member of this family is the two-spotted spider mite *Tetranychus urticae* which is a major pest in a wide range of crops worldwide ((Van Leeuwen *et al.* 2010a), Fig. I.1*A*). It is one of the most polyphagous arthropods known to man (Migeon and Dorkeld 2010) and is also notorious for its ability to rapidly develop resistance to pesticides (Van Leeuwen *et al.* 2010a). The Eriophyoidea are the second most important phytophagous mite species complex (Van Leeuwen *et al.* 2010b). In most cases, these worm-like four legged mites have a restricted host plant range and the symptoms of their feeding vary from russeting to complex gall formation ((Skoracka *et al.* 2010), Fig. I.1*F*).

The control of phytophagous mites, particularly spider mites, can often also be achieved by biological control. Predatory phytoseiid mites (Parasitiformes: Mesostigmata: Phytoseiidae, Fig. I.1*G*) are widely used for this purpose (Gerson *et al.* 2003). Three species of this group belong to the ten most important biological control agents used in augmentive biological control (van Lenteren 2012).

Among agriculturally important parasitic mites, the honeybee mite *Varroa destructor* (Parasitiformes: Mesostigmata: Varroidea, Fig. I.1*E*) is considered as one of the greatest threats to apiculture, with total damage ranging in billions of dollars (Cook *et al.* 2007; Garbian *et al.* 2012). Besides feeding on the hemolymph of developing and mature honeybees, this obligate parasite is also an important vector of bee viruses (Garbian *et al.* 2012). Next, the chicken mite *Dermanyssus gallinae* (Parasitiformes: Mesostigmata: Dermanyssidae) is an important ectoparasite of poultry. This blood feeding mite is

responsible for egg downgrading and aenemia in chickens, reducing their welfare and causing mortality. The financial loss due to the chicken mite in the Netherlands has been estimated at 11 million euro/year (Emous *et al.* 2005; Mul *et al.* 2009; Sparagano *et al.* 2009). However, the most famous blood feeding parasitic Acari are ticks (Parasitiformes: Ixodida, Fig. I.1*D*). During their blood meal, they can cause infection and irritation of the skin and transmit disease. Of all arthropod vector groups, ticks are able to transmit the greatest variety of pathogenic microorganisms, protozoans, rickettsiae, spirochaetes and viruses. They are among the most important vectors of diseases affecting livestock, humans and companion animals. Their global economic importance is especially high for livestock (e.g. *Rhipicephalus microplus*), but there is also a significant impact on public health in the northern hemisphere, mainly due to the fact that some tick species (e.g. *Ixodes* sp.) are able to transmit *Borrelia burgdorferi*, the causing agent of Lyme disease. For example, in the period 2000-2010, more than 250 000 confirmed cases of Lyme disease were reported in the USA (Jongejan and Uilenberg 2004; Peter *et al.* 2005; Nicholson *et al.* 2009; Embers and Narasimhan 2013).

Besides ticks, also house dust mites (Fig. I.1*B*), itch mites (Acariformes: Sarcoptiformes: Astigmata) and trombiculid mites (Acariformes: Sarcoptiformes: Prostigmata) are of medical importance. About 40 years ago, house dust mites were recognized as one of the major sources of allergens in house dust (Voorhorst *et al.* 1967). In sensitive persons, these allergens are able to cause asthma, dermatitis or rhinitis (Arlian and Platts-Mills 2001). In countries with a temperate climate, 6 to 35 per cent of the population is sensitive to house dust mite-derived allergens (Janson *et al.* 2001). The itch mite, *Sarcoptes scabiei*, is the causing agent of scabies, a common but neglected skin disease. The female itch mites burrow in the epidermis of their hosts and feed on the cells, while they deposit their eggs and faecal pellets. Worldwide it has been estimated that 300 million people, mainly in empoverished communities, are currently infested with the scabies mite (Hengge *et al.* 2006; Mounsey *et al.* 2013). Finally, larvae (chiggers) of several trombiculid mites of the Pacific region, Southern Asia and Australia are known as vectors of *Orientia tsutsugamushi*. This rickettsia is transmitted to humans when they are bitten by trombiculid larvae and causes a human infectious disease named scrub typhus. There are about one million cases of scrub typhus a year and one billion people are believed to be at risk (Kuo *et al.* 2011).

2 Acari genomics

2.1 Introduction

In 1976 Walter Fiers and his colleagues deciphered, for the first time, the complete sequence of a genome (Fiers *et al.* 1976). One year later Sanger *et al.* (1977) developed a new and reliable method for DNA–sequencing, the chain termination method. These landmark discoveries marked the beginning of a new era in genetics: genomics. Genomics is the study of structure, function and evolution of genomes (Heckel 2003; Hunter and Chittarankan 2008; Chilana *et al.* 2012).

Only four years after Sanger's invention the complete human mitochondrial (mt) genome was determined (Anderson *et al.* 1981) while in 1998 the first complete nuclear genome of a multicellular organism, the nematode *Caenorhabditis elegans*, was sequenced (*C. elegans* Sequencing Consortium 1998). In recent years, the DNA sequencing boom has made it economically feasible to obtain the complete nuclear genome sequence of many multicellular organisms (Fig. I.3). Remarkably, among the sequenced nuclear non-human animal genomes that have been published (status June 2012), the phylum of the Arthropoda ranks the highest (Chilana *et al.* 2012; Song and Wang 2013) (Fig. I3).

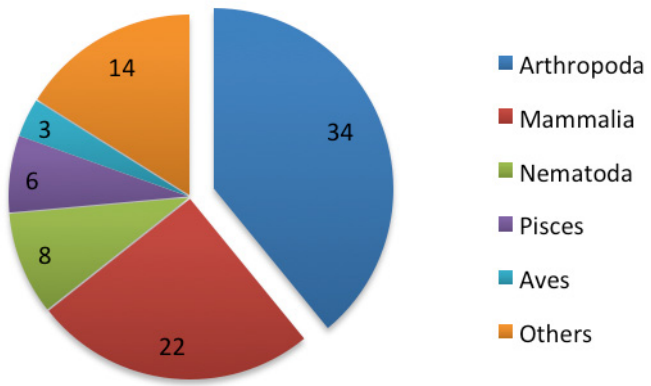


Figure I.3 – Pie chart of different non-human animal taxa with publications on their genome projects (drafted from Song *et al.* 2013).

2.2 Arthropod genomics

Arthropod cells, like most eukaryotic cells, contain a genome both in the nucleus and the mitochondria. Both genomes are the subject of study in genomics. Hence, an overview of available arthropod mt and nuclear genomes will be presented in the following two subsections. As a transcriptome can be considered as the initial product of genome expression, a survey of arthropod transcriptomes will be briefly discussed in the last subsection.

2.2.1 Arthropod mitochondrial genomes

Currently, 550 arthropod mt genome sequences have been deposited in the NCBI database (MetaMIGA database (Feijao *et al.* 2006), status 15 February 2013). Most of these mt sequences (365) were derived from species belonging to the class Insecta. Ninety-seven and (only) 64 belong to species from the subphyla Crustacea and Chelicerata, respectively. With a few exceptions, typical arthropod mt genomes have the same characteristics as most metazoan mt genomes. They are circular, between 13 and 20 kb in length, contain a coding region with 37 genes (22 tRNAs, 2 rRNAs (12S and 16S rRNA) and 13 intronless protein coding genes (PCG, *cox1*, *cox2*, *cox3*, *cytB*, *atp6*, *atp8*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* and *nad6*)) and a relatively small non-coding region. The latter mostly fulfills a role in transcription and replication (Wolstenholme 1992; Boore 1999; Taanman 1999; Gissi et al. 2008; Bernt et al. 2013). In addition, arthropod mt genomes are highly A+T biased (Simon *et al.* 1994), with the currently available mt genomes having an average A+T content of 74,2±5,5% (MetaMIGA database (Feijao *et al.* 2006), status 15 February 2013). Like most animals, arthropod mt DNA exhibits a bias between the two DNA strands in the composition of A versus T and C versus G (Hassanin *et al.* 2005). This strand bias can be measured by calculating GC- and AT-skews ((G%-C%)/(G%+C%) and (A%-T%)/(A%+T%), respectively (Perna and Kocher 1995; Torricelli *et al.* 2010). In arthropods there is usually more A than T (positive AT-skew) and more C than G (negative GC-skew) on the majority (J) strand, which is the strand coding for most of the mt genes. Logically, the minority strand (N) contains less genes than the J-strand and has opposite GC- and AT-skews. The N-strand of most arthropod mt genomes corresponds, because it contains an excess of heavier nucleotides (G > C, T > A), to the heavy strand (H) of mammalian mt genomes, while the J-strand corresponds to the light strand (L) (Simon *et al.* 1994; Wei *et al.* 2010). The clear strand bias in nucleotide composition in metazoan mt genomes is probably due to asymmetric patterns of mutations during transcription and replication when one strand remains in a single stranded state,

making it more vulnerable to DNA damage (see Fig. I.4 for a replication model of animal mtDNA).

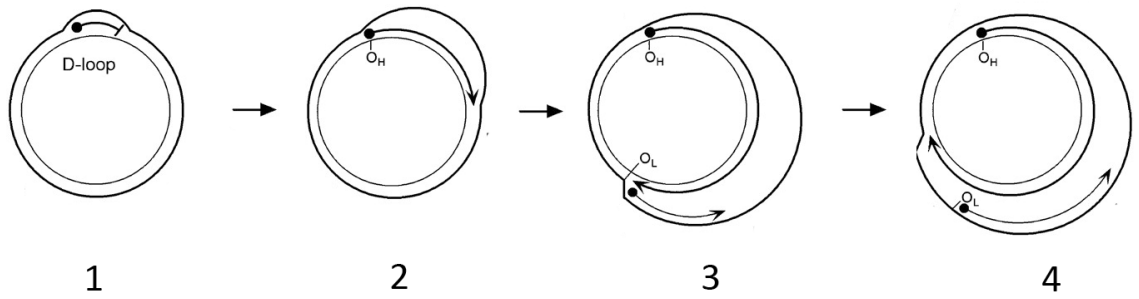


Fig. I.4 – The strand-displacement model of animal mtDNA replication (drafted from Brown *et al.* 2005)

The strand-displacement model of animal mtDNA replication is the most widely accepted model of animal mtDNA replication (Wei *et al*. 2010). Replication starts at the H-strand replication origin (O_H), located on the L-strand in the main non coding-region of the mtDNA. This non-coding region is also called control region or D-loop (1). A triple stranded structure is formed because of the elongation of the nascent H-strand, displacing the parental H-strand (2). This proceeds until the L-strand replication (O_L) origin is exposed (3), with subsequent synthesis of the new L-strand in the opposite direction (3-4). As a consequence of its single stranded state the H-strand is supposed to be more exposed to mutations than the L-strand (Brown *et al.* 2005; Hassanin *et al.* 2005*)*.

Furthermore, in some arthropod species the mt genetic code deviates from the standard invertebrate mt code, with the AGG codon coding for lysine instead of serine (Abascal *et al.* 2012). The inheritance of mt DNA is strictly maternal, with the exception of some bivalves which have a doubly uniparental inheritance (Breton *et al.* 2007). With a few exceptions (see 2.3.1 Acari mt genomes) the arthropod mt gene order is highly conserved, and similar to the mt gene arrangement of the horseshoe crab, *Limulus polyphemus*, which is generally considered as the representative ground pattern for arthropod mt genomes (Fig. I.5)(Lavrov *et al.* 2000; Stach *et al.* 2010). Similar to humans, transfers of DNA from the mitochondria to the nucleus, also called NUMTs (nuclear copies of mt origin), occur in arthropods, with the honeybee nuclear genome having an exceptionally high density of NUMTs (> 1bp NUMTs/kb) (Pamilo *et al.* 2007; Black and Bernhardt 2009; Hazkani-Covo *et al.* 2010). Remarkably, in some cases arthropod mt genomes extremely deviated from the canonical mt genomes. In the case of the terrestrial isopod *Armadillidium vulgare*, the mt genome is 42 kb long consisting of two molecules co-occurring in mitochondria: a circular approximately 28 kb dimer formed by two approximately 14 kb monomers fused in opposite polarities, and a

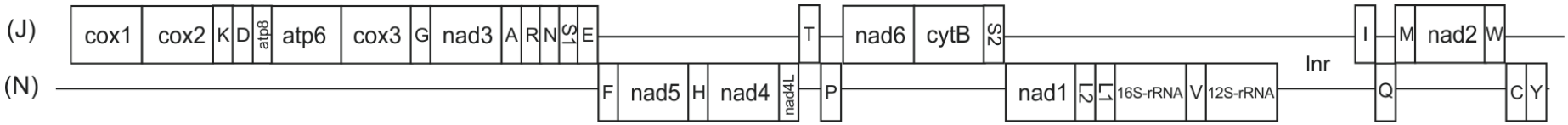


Figure I.5 – Mitochondrial genome of the horseshoe crab *Limulus polyphemus*.

Graphical linearisation of the *L. polyphemus* mt genome according to Fahrein *et al.* (2007) is presented . Gene sizes are not drawn to scale. J stands for majority and N for minority strand. Protein coding and rRNA genes are abbreviated as in the abbreviations section. tRNA genes are abbreviated using the one-letter amino acid code, with L_1=CUN; L_2=UUR; S_1= AGN; S_2=UCN.

linear approximately 14 kb monomer (Marcadé *et al.* 2007). The mt genomes of the human body louse and head louse, on the other hand, seems to be scattered over 20 minichromosomes with each minichromosome being 3 to 4 kb in size, and having 1 to 3 genes and a control region (Shao *et al.* 2012). Finally, the mt genome of the booklouse *Liposcelis bostrychophila* is reported to be spread over 2 mt chromosomes, each being about 8 kb in size and each containing about half of the set of mt genes (Wei *et al.* 2012)

2.2.2 Arthropod nuclear genomes

2.2.2.1 Sequenced arthropod nuclear genomes

As of February 2013, 61 arthropod nuclear genomes have been sequenced and deposited in the NCBI-database (http://www.ncbi.nlm.nih.gov/genome/browse/) (Table I.2). The number of available arthropod nuclear genomes is clearly skewed towards insects. The first published arthropod nuclear genome was that of the fruit fly *Drosophila melanogaster*, which became the most-studied eukaryotic genome and the most-prominent model organism in molecular biology ((Adams *et al.* 2000), Table I.2). The remaining sequenced arthropod genomes mainly belong to the insect orders of the Diptera (29), Hymenoptera (12) and Lepidoptera (5). To date, only one crustacean and five chelicerate genome(s) have been sequenced. However, it can be expected that in the near future these numbers will quickly change as two years ago a project named *i5K* has been launched. This international consortium aims to sequence 5000 arthropod genomes in the next 5 years (Robinson *et al.* 2011). In addition, many other individual arthropod genome projects (>100) have been started but have not been completed yet (e.g. see http://www.hgsc.bcm.tmc.edu/content/arthropod-sequencing#ongoing (Baylor College of Medicine), http://www.genomics.cn/en/navigation/show_navigation?nid=5680 (Beijing Genomics Institute) and http://www.genomesonline.org).

2.2.2.2 Nuclear genome sequencing

To date, the majority of arthropod nuclear genomes have been sequenced using the shotgun Sanger sequencing method. In this method, high molecular weight genomic DNA is isolated, randomly sheared and ligated into plasmid vectors. Subsequently, the plasmid DNA is isolated and sequenced using automated Sanger sequencing (Palli *et al.* 2012) (Fig. I.6). The last 5 years, however, next-generation sequencing (NGS) technologies (Roche /454, Illumina, SOLiD) have entered the market. These new sequencing platforms are quite diverse in template preparation and sequencing biochemistry but have in common that they have a much higher throughput than Sanger sequencing (for a review of NGS technologies, see

Shendure and Ji 2008 and Metzker 2010). Using NGS technologies, sequencing of new genomes will be completed more quickly and at a more affordable cost compared to the insect genomes sequenced to date (Metzker 2010; Palli *et al.* 2012). For example, about 10 years ago the sequencing cost of the honey bee genome (236 Mb large) was about 9 million dollars (Hackett 2011) while the 484 Mb large nuclear genome of the fire ant *Solenopsis invicta* was sequenced using a combination of Roche/454 and Illumina technologies at a cost of "only" 230,000 dollar (Wurm *et al.* 2011).

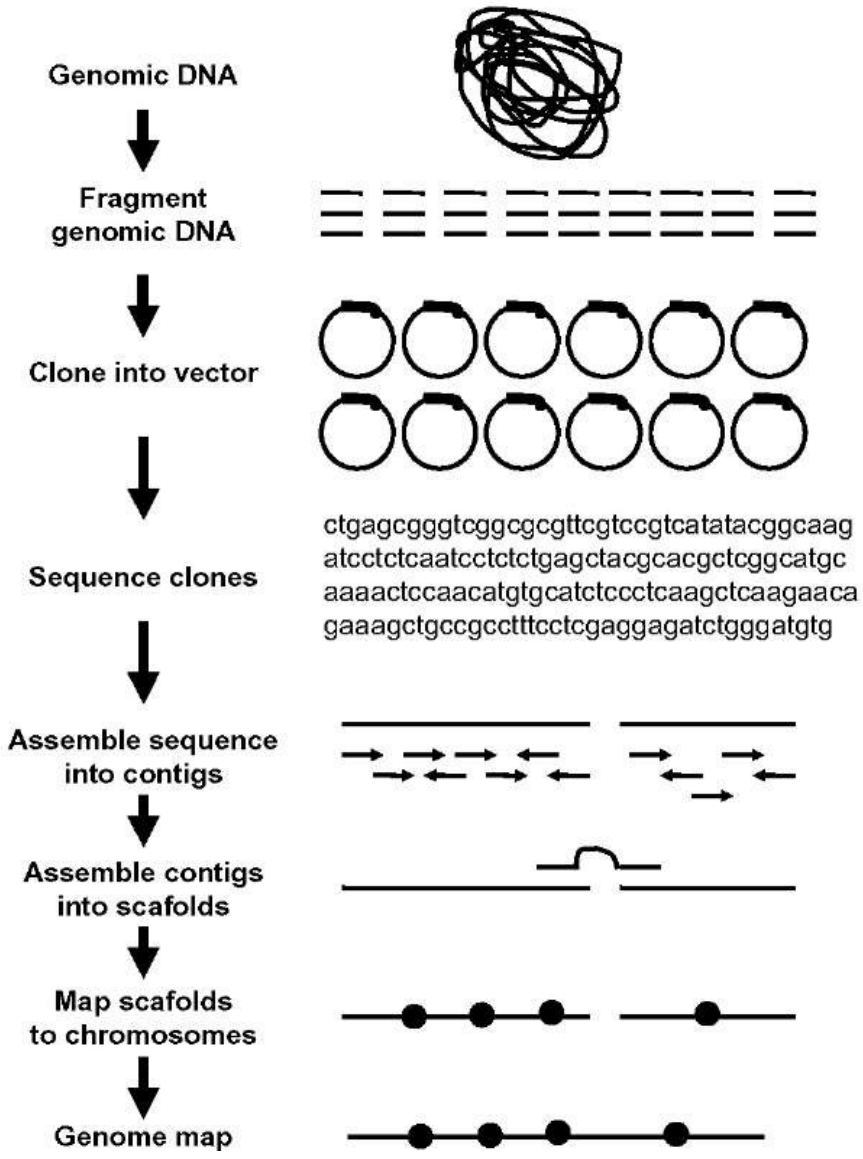


Figure I.6 - Sanger shotgun sequencing method (drafted from Palli *et al.* 2012)

Table I.2 - Arthropod genomes that have been sequenced and deposited in the NCBI-database (status 15 February 2013)

Species	Importance	Taxonomy	Genome size (Mb)	Number of genes	Reference*
Tetranychus urticae	crop pest	Chelicerata: Acari	90	18,414	Grbic *et al.* 2011
Ixodes scapularis	parasite of cattle/disease vector	Chelicerata: Acari	1,896	7,112	-
Metaseiulus occidentalis	predatory mite	Chelicerata: Acari	152	12,455	-
Rhipicephalus microplus	parasite of cattle/disease vector	Chelicerata: Acari	145	-	-
Varroa destructor	bee parasite	Chelicerata: Acari	313	-	-
Daphnia pulex	model organism	Crustacea: Branchiopoda	200	30,907	Colbourne *et al.* 2011
Lepeophtheirus salmonis	salmon louse	Crustacea: Copepoda	600	-	-
Aedes aegypti	disease vector	Insecta: Insecta	1,380	15,419	Arensburger *et al.* 2010
Anopheles sp. (3)	disease vector	Insecta: Insecta	114-278	11,708-12,457	Arensburger *et al.* 2010,-
Culex pipiens quinquefasciatus	disease vector	Insecta: Insecta	579	18,883	Arensburger *et al.* 2010
Drosophila melanogaster	fruit fly - model organism	Insecta: Diptera	180	15,969	Adams *et al.* 2000, Flybase Release Notes 2013
Other *Drosophila* sp. (20)	model organism	Insecta: Diptera	130-236	15,179-17,573	a,-
Lutzomyia longipalpis	disease vector	Insecta: Diptera	143	-	-
Mayetiola destructor	crop pest	Insecta: Diptera	153	-	-
Phlebotomus papatasi	disease vector	Insecta: Diptera	345	-	-
Acyrthosiphon pisum	crop pest	Insecta: Hemiptera	517	36,604	Richards *et al.* 2010
Rhodnius prolixus	disease vector	Insecta: Hemiptera	562	-	-
Linepithema humile	Argentine ant, household pest	Insecta: Hymenoptera	251	16,123	Smith *et al.* 2011a
Other ant species (6)	Miscellaneous	Insecta: Hymenoptera	240-484	16,331-18,564	b
Apis florea	pollinator	Insecta: Hymenoptera	213	-	-
Apis mellifera	pollinator	Insecta: Hymenoptera	236	10,157	Weinstock *et al.* 2006
Bombus terrestris	pollinator	Insecta: Hymenoptera	249	10,178	-
Megachile rotundata	solitary bee, pollinator	Insecta: Hymenoptera	266	-	-
Nasonia sp. (3)	model for parasitoid genetics	Insecta: Hymenoptera	180-295	17,279	Werren *et al.* 2010,-
Tribolium castaneum	stored grain pest	Insecta: Coleoptera	160	16,404	
Plutella xylostella	agricultural pest	Insecta: Lepidoptera	394	18,071	You *et al.* 2013
Bombyx mori	silkworm, model organism	Insecta: Lepidoptera	429	18,510	Mita *et al.* 2004, Xia *et al.* 2004
Danaus plexippus	Monarch butterfly	Insecta: Lepidoptera	273	16,866	Zhan *et al.* 2011
Heliconius melpomene	model for the evolution of mimicry	Insecta: Lepidoptera	269	12,669	Dasmahapatra *et al.* 2012
Manduca sexta	crop pest - model organism	Insecta: Lepidoptera	400	-	-
Pediculus humanus	disease vector	Insecta: Phtiraptera	108-110	10,773	Kirkness *et al.* 2010
Mengenilla moldrzyki	important position in insect phylogeny	Insecta: Strepsitera	165	16,772	Niehuis *et al.* 2012
Strigamia maritima	important position in arthropod phylogeny	Myriapoda	174	-	-

* genome sizes and gene numbers of published genomes were derived from the cited reference while not (yet) published/analyzed genomes (-) were derived from the NCBI genome database (https://www.ncbi.nlm.nih.gov/genome/browse/)

a Richards *et al.* 2005, Clark *et al.* 2007, Flybase Release Notes 2013_01 (http://flybase.org/static_pages/docs/release_notes.html)

b Bonasio *et al.* 2010, Suen *et al.* 2010, Nygaard *et al.* 2011, Smith *et al.* 2011b, Wurm *et al.* 2011

2.2.3 Arthropod transcriptomes

The initial product of genome expression is the transcriptome, a collection of RNA transcripts from which biological information is required by the cell at a particular time (Brown 2002). Transcriptome sequencing has several advantages compared to genome sequencing like the smaller size of the transcriptome (resulting in lowering sequencing costs), fewer repetitive elements (which reduces analytical burden during post-sequencing assembly) and the possibility to detect rare transcripts with regulatory roles (Riesgo *et al.* 2012). To obtain an overview of available transcriptomes for the various arthropod subphyla, classes and orders, the NCBI-database was mined for the presence of arthropod sequences in the expressed sequence tags (ESTs) database using the Taxonomy browser. This database contains first-pass single-reads of cDNA clones. Although many arthropod transcriptomes are also deposited under the form of sequence read archives (SRA) in the NCBI database, these were not included in our analysis. SRAs can also contain data from whole genome sequencing projects and it was not possible to filter these data from the transcriptome data.

Table I.3 - Number of arthropod ESTs deposited in the NCBI-database (status February 15, 2013)

Arthropod group	Number of ESTs
Insecta	**5,312,207**
Diptera	2,252,438
Lepidoptera	1,244,367
Hemiptera	688,516
Hymenoptera	461,573
Coleoptera	345,178
Orthoptera	108,134
Other insect Orders	212,001
Chelicerata	**452,947**
Acari	397,595
Parasitiformes	298,610
Ixodida	298,610
Mesostigmata	0
Acariformes	98,985
Astigmata	18,130
Prostigmata	80,855
T. urticae	80,855
Araneae	38,105
Other chelicerate orders	17,247
Crustacea	**902,251**
Other arthropod groups	**51,059**
Arthropoda	**6,718,464**

The number of available arthropod ESTs is clearly biased towards the Insecta (Table I.3). About 80% of all available arthropods EST belong to species from this class. Not surprisingly, orders containing many agriculturally and medically important species like Diptera (flies), Lepidoptera (butterflies and moths), Hemiptera (aphids, cicadas and true bugs) and Coleoptera (beetles) have the highest numbers of available ESTs within the Insecta. Next, 16 % of available arthropod ESTs belong to crustacean species while less than 10% belongs to members of the Chelicerata. Within the chelicerate subclass of the Acari, the number of available ESTs is skewed towards the Ixodida (ticks), with 75% of all Acari ESTs, while only 25% belong to species of the Acariformes, the most diverse and species-rich group of the Acari (see 1.2 Classification). Remarkably, 80% of the number of available acariform ESTs belong to only one species, the spider mite *T. urticae* of which the genome has been recently sequenced (Grbić *et al.* 2011).

Similar to the overview of arthropod genomes (Table I.2), Table I.3 reflects only a temporary status, since many arthropod transcriptomes are being sequenced or are proposed for sequencing. In analogy with arthropod genomes, NGS technologies allow to sequence transcriptomes at higher speed and lower cost. For example, a project named 1KITE aiming to sequence transcriptomes of 1000 insects belonging to various insect orders was started in September 2011 and should be nearly completed at the time of writing (http://www.1kite.org/).

2.3 Acari genomics

2.3.1 Acari mitochondrial genomes

To date the mt genomes of 36 Acari species have been completely sequenced (Table I.4). Fourteen Acari mt genomes belong to species of the Acariformes and 22 to members of the Parasitiformes. Acari mt genomes have an average length of 14,872±1928 bp and, like most arthropod mt genomes, a high A+T content (76±4,52 %). The average AT-skew of Acari mt genomes is 0.006±0.099 ranging from -0.253 in the sarcoptiform mite *Dermatophagoides farinae* to 0.274 in the trombidiform mite *Unionicola parkeri*. The average GC-skew of Acari mt genomes is -0.141±0.174 with most Acari mt genomes having a negative GC-skew, similar to what is found for most arthropod mt genomes (see 2.2.1 Arthropod mt genomes). Six mite species have a reversal of both the GC- and AT-skew compared to those of most arthropod mt genomes: *Dermatophagoides pteronyssinus*, *D. farinae*, *Phytoseiulus persimilis*, *Panonychus citri*, *Panonychus ulmi* and *Varroa destructor* (Table I.4). This reversal could be

the result of an inversion of the mt control region (see Chapter II and III) (Hassanin *et al.* 2005; Yuan *et al.* 2010). An inversion or "strand swap" of the control region is expected to produce a global reversal of asymmetrical mutational constraints during mtDNA replication (see 2.1.1 Arthropod mitochondrial genomes), resulting with time in a complete reversal of strand compositional bias (Hassanin *et al.* 2005, Wei *et al.* 2010).
Remarkably, some Acari mt genomes show alterations on the typical gene content of animal mt genomes (see 2.2.1 Arthropod mt genomes). The mt genome of the chigger mite *Leptotrombidium pallidum* has a duplication of the 16s-RNA gene while the mt genome of *S. magnus* lacks 16 tRNAs (Shao *et al*. 2005a; Domes *et al*. 2008). Strikingly, the mt genome of *M. occidentalis* is 24,961 bp large, possesses a 9,921 bp duplicated region and lacks *nad3* and *nad6* (Jeyaprakash and Hoy 2007). Nevertheless, a reevaluation of this genome revealed that this genome is smaller than previously reported and that the protein coding gene *nad3* is in fact not lacking (see Chapter III).

In contrast to most arthropod mt genomes, a great variation is found among Acari mt gene orders. The mt gene order of the tick *Nuttalliella namaqua* (belonging to the monotypic family Nuttalliellidae), the prostriate hard ticks and the soft ticks is identical to the ancestral arthropod mt ground pattern, while the mt gene order of metastriate hard ticks exhibits a small deviation (translocation of a block of genes (*nad1* – tRNA of glutamine)) on this pattern (Black and Roehrdanz 1998; Shao *et al.* 2004; Burger *et al.* 2012; Mans *et al.* 2012). Members of the Mesostigmata and Acariformes on the other hand exhibit highly rearranged mt gene orders (see Chapter II and III) (Navajas *et al.* 2002; Yuan *et al.* 2010; Edwards *et al.* 2011). All together, this indicates that Acari mt gene arrangements have probably been independently derived and limits the use of mt gene orders in inferring phylogenetic relationships between Acari superorders (see Chapter II) (Fahrein *et al.* 2007; Yuan *et al.* 2010). Other mt genome characters that have been extensively studied in Acari are tRNAs and rRNAs. The average size of tRNAs of Acariformes is significantly shorter than those of Parasitiformes and those of the horseshoe crab and mt genome model organism *Limulus polyphemus*. Furthermore, acariform tRNAs are atypical and lack either the T or D- arm, deviating from the typical cloverleaf tRNA structure (see Chapter II and III) (Shao *et al*. 2005a, Shao *et al*. 2006; Domes *et al*. 2008; Klimov and Oconnor 2009; Yuan *et al*. 2010). Interestingly, both rRNAs (12S and 16S) of Acariformes are also considerably shorter ("minimal rRNAs") than those of Parasitiformes and most other metazoan species (Table I.4,

Table I.4 - Sequenced Acari mt genomes (as derived from the MetaMIGA database, status 15 February 2013)

Refseq number	Species	Taxonomy[1]		Length (bp)	A+T content (%)	AT-skew	GC-skew	12S-rRNA length (bp)	16S-rRNA length (bp)
NC_013184	*Dermatophagoides farinae*	A: Sarcoptiformes: Astigmata: Pyroglyphidae		14,266	71.4	-0.253	0.231	661	1,017
NC_012218	*Dermatophagoides pteronyssinus*	A: Sarcoptiformes: Astigmata: Pyroglyphidae		14,203	72.6	-0.199	0.194	665	1,078
NC_011574	*Steganacarus magnus*	A: Sarcoptiformes: Oribatida: Steganacaridae		13,818	74.6	-0.020	-0.037	609	992
NC_007601	*Leptotrombidium akamushi*	A: Trombidiformes: Prostigmata: Trombiculidae		13,698	67.5	-0.016	-0.075	596	1,026
NC_007600	*Leptotrombidium deliense*	A: Trombidiformes: Prostigmata: Trombiculidae		13,731	70.0	-0.017	-0.058	602	1,023
NC_007177	*Leptotrombidium pallidum*	A: Trombidiformes: Prostigmata: Trombiculidae		16,779	71.0	-0.031	-0.044	601	1,007
NC_010595	*Walchia hayashii*	A: Trombidiformes: Prostigmata: Trombiculidae		14,857	73.0	0.264	-0.305	625	1,045
NC_011036	*Unionicola foili*	A: Trombidiformes: Prostigmata: Unionicolidae		14,738	73.0	0.201	-0.279	649	1,016
NC_014683	*Unionicola parkeri*	A: Trombidiformes: Prostigmata: Unionicolidae		14,734	72.8	0.279	-0.272	653	1,017
NC_010596	*Ascoschoengastia* sp, TATW	A: Trombidiformes: Prostigmata: Trombiculidae		16,067	70.1	0.015	-0.049	680	1,047
NC_014347	*Panonychus citri*	A: Trombidiformes: Tetranychidae		13,075	85.5	-0.075	0.033	651	980
NC_012571	*Panonychus ulmi*	A: Trombidiformes: Tetranychidae		13,115	85.6	-0.059	0.005	659	982
NC_014399	*Tetranychus cinnabarinus*	A: Trombidiformes: Tetranychidae		13,092	84.4	0.031	-0.028	644	991
NC_010526	*Tetranychus urticae*	A: Trombidiformes: Tetranychidae		13,103	84.3	0.026	-0.016	646	991
NC_019642	*Argas africolumbae*	P: Ixodida: Argasidae		14,440	73.4	0.053	-0.319	695	1,198
NC_005291	*Carios capensis*	P: Ixodida: Argasidae		14,418	73.5	0.036	-0.374	695	1,225
NC_004357	*Ornithodoros moubata*	P: Ixodida: Argasidae		14,398	72.3	0.067	-0.379	686	1,212
NC_005820	*Ornithodoros porcinus*	P: Ixodida: Argasidae		14,378	71.0	0.059	-0.355	691	1,207
NC_017758	*Amblyomma elaphense*	P: Ixodida: Ixodidae: Metastriata		14,627	80.5	-0.032	-0.137	696	1,152
NC_017745	*Amblyomma sphenodonti*	P: Ixodida: Ixodidae: Metastriata		14,772	77.8	-0.028	-0.124	712	1,204
NC_005963	*Amblyomma triguttatum*	P: Ixodida: Ixodidae: Metastriata		14,740	78.4	-0.022	-0.133	693	1,199
NC_017759	*Aponomma fimbriatum*	P: Ixodida: Ixodidae: Metastriata		14,705	77.7	-0.019	-0.157	694	1,211
NC_017756	*Bothriocroton concolor*	P: Ixodida: Ixodidae: Metastriata		14,809	75.1	-0.022	-0.127	698	1,222
NC_017757	*Bothriocroton undatum*	P: Ixodida: Ixodidae: Metastriata		14,769	76.9	-0.038	-0.097	706	1,213
NC_005292	*Haemaphysalis flava*	P: Ixodida: Ixodidae: Metastriata		14,686	76.9	-0.018	-0.116	699	1,196
NC_002010	*Ixodes hexagonus*	P: Ixodida: Ixodidae; Prostriata		14,539	72.7	0.033	-0.366	705	1,287
NC_005293	*Ixodes holocyclus*	P: Ixodida: Ixodidae; Prostriata		15,007	77.4	-0.013	-0.254	716	1,214
NC_004370	*Ixodes persulcatus*	P: Ixodida: Ixodidae; Prostriata		14,539	77.3	-0.024	-0.269	720	1,206
NC_018369	*Ixodes ricinus*	P: Ixodida: Ixodidae; Prostriata		14,566	78.7	-0.024	-0.262	718	1,296
NC_006078	*Ixodes uriae*	P: Ixodida: Ixodidae; Prostriata		15,053	74.8	0.007	-0.328	712	1,210
NC_002074	*Rhipicephalus sanguineus*	P: Ixodida: Ixodidae: Metastriata		14,710	78.0	-0.034	-0.098	687	1,190
NC_019663	*Nuttalliella namaqua*	P: Ixodida: Nutalliellidae		14,425	78.6	0.035	-0.290	733	1,164
NC_013474	*Stylochyrus rarior*	P: Mesostigmata: Ologamasidae		14,900	72.7	0.054	-0.305	653	1,200
NC_009093	*Metaseiulus occidentalis*	P: Mesostigmata: Phytoseiidae		24,961	76.0	0.095	-0.291	742	1,192
NC_014049	*Phytoseiulus persimilis*	P: Mesostigmata: Phytoseiidae		16,199	79.8	-0.062	0.222	711	1,199
NC_004454	*Varroa destructor*	P: Mesostigmata: Varroidae		16,476	80.0	-0.021	0.178	726	1,149
			*	14,872±1,928	76.0± 4.5	0.006±0,099	-0.141±0.174	678±39	1,132±100
			**	14,234±1,131	75.4± 6.5	0.010±0.153	-0.050±0.156	639±27	1,015±27
			***	15,278±2,227	76.3± 2.8	0.003±0.042	-0.199±0.162	704±19	1,206±33

[1] A: Acariformes; P: Parasitiformes
* calculated average of mt genome values of all Acari
** calculated average of mt genome values of Acariformes
***calculated average of mt genome values of Parasitiformes

see Chapter II) (Yuan *et al.* 2010; Klimov and Knowles 2011). Wolstenholme *et al.* (1987) previously suggested that truncated rRNAs and atypical tRNAs evolved in a concerted fashion in nematodes. Whether this hypothesis is true and, by extension, also applies for acariform mites is intriguing and should be confirmed by additional experiments.

2.3.1 Acari nuclear genomes

2.3.1.1 General

To date five Acari genomes have been (partially) sequenced, of which only the *T. urticae* genome has been reported to be complete (Table I.5) (Grbić *et al.* 2011). The genome sequence of *M. occidentalis* is probably complete with an assembly of 152 Mb and estimated genome size of 88-90 Mb (Table I.5) (Hoy 2009; Jeyaprakash and Hoy 2009a) while the genomes of *V. destructor*, *R. microplus* and *I. scapularis* have only partially been sequenced (Cornman *et al.* 2010; Guerrero *et al.* 2010; Hill 2010). The genome of *V. destructor* has an estimated size of 565±3 Mb, while the assembly summed to only 318 Mb of sequence (Cornman *et al.* 2010). *R. microplus* and *I. scapularis* have an estimated genome size of 7.1 and 2.1 Gb while only 145 and 1896 Mb of sequence have been assembled to date (Table I.5) (Ullmann *et al.* 2005). The large size of the latter 2 genomes has been attributed to a significant amount of non-coding repetitive DNA in these genomes (Van Zee *et al.* 2007). Due to its high repetitive nature and large size, the *R. microplus* genome will, with present technologies and costs, not be completed in the near future (Guerrero *et al.* 2010) while in 2010 a new strategy was proposed to sequence the remaining part of the *I. scapularis* genome (Hill 2010).

Next to these sequenced genomes, other mite and tick genomes have been recently proposed to be sequenced or are being sequenced (Table I.5). As one can expect, most of these candidate genomes mainly belong to agriculturally, veterinary or medically important species (Table I.5). Genomes belonging to mites of the diverse and phylogenetically important order of the Oribatida have not yet been proposed, probably due to their minor economical importance. Surprisingly, genome projects regarding the medically important house dust mites have also, to our knowledge, not yet been started. Interestingly, in contrast to Ixodida, estimated genome sizes of Acariformes (Prostigmata and Astigmata) are among the smallest of all arthropods (Table I.2 and I.5) (Hanrahan and Johnston 2011), making them ideal candidates to be sequenced at low-cost.

Table I.5 - An overview of sequenced and future (to be) sequenced Acari genomes

Species	Taxonomy	Common name	Importance*	Genome Assemby Size (Mb)[a]	Estimated Genome Size (Mb)	Number of genes	Reference(s)
Sequenced Acari genomes							
Tetranychus urticae	Prostigmata: Tetranychidae	two-spotted spider mite	A	90	90	18,414	Grbić *et al.* 2011[b]
Metaseiulus occidentalis	Mesostigmata: Phytoseiidae	western predatory mite	A	152	88-90	12,455	Hoy 2009[c]
Varroa destructor	Mesostigmata: Varroidae	honey bee mite	A	313	562-568		Cornman *et al.* 2010
Ixodes scapularis	Ixodida: Ixodidae	deer tick	V, M	1,896	2,100	7,112	Hill 2010[c]
Rhipicephalus microplus	Ixodida: Ixodidae	southern cattle tick	V	145	7,100		Guerrero *et al.* 2006[d]
Proposed or being sequenced							
Sarcoptes scabiei	Astigmata: Sarcoptidae	itch mite	M		89-103		i5K[e,f]
Ornithodoros sp. (2)	Ixodida: Argasidae	(soft ticks)	V		1,090±120 (*O. turicata*)		i5K[e], Hill 2010[g]
Amblyomma sp. (3)	Ixodida: Ixodidae	lone star tick (*A. americanum)*	V		3,108±27 (*A. americanum)*		i5K[e], Hill 2010[g]
Dermacentor variabilis	Ixodida: Ixodidae	American dog tick/Wood tick	V		2,862±79		i5K[e], Hill 2010[g]
Hyalomma dromedarii	Ixodida: Ixodidae	camel tick	V				i5K[e]
Other *Ixodes* sp. (3)	Ixodida: Ixodidae	(hard ticks)	V, M				i5K[e], Hill 2010
Dermanyssus gallinae	Mesostigmata: Dermanyssidae	poultry red mite	V				i5K[e]
Aceria tosichella	Prostigmata: Eriophyidae	wheat curl mite	A				i5K[e]
Aculops lycopersici	Prostigmata: Eriophyidae	tomato rust mite	A				[h]
Phytonemus pallidus	Prostigmata: Tarsonemidae	cyclamen mite	A				i5K[e]
Brevipalpus sp. (3)	Prostigmata: Tenuipalpidae	false spider mites/flat mites	A				i5K[e]
Bryobia praetiosa	Prostigmata: Tetranychidae	clover mite	A				i5K[e]
Panonychus citri	Prostigmata: Tetranychidae	citrus red mite/ citrus fruit mite	A				i5K[e]
Tetranychus cinnabarinus	Prostigmata: Tetranychidae	carmine spider mite	A				i5K[e]
Tetranychus evansi	Prostigmata: Tetranychidae	tomato red spider mite	A				Rombauts *et al.* 2013
Tetranychus lintearius	Prostigmata: Tetranychidae	gorse spider mite	A				Rombauts *et al.* 2013
Leptotrombidium deliense	Prostigmata: Trombiculidae	chigger mite	M				Hill 2010

* A= Agricultural, V= Veterinary, M=Medical

[a] except for *T. urticae* genome assembly sizes were derived from the NCBI-database

[b] genomes assembly size and number of genes for *T. urticae* were derived from Grbic *et al.* (2011) while estimated genome size was derived from Hanrahan and Johnston 2011

[c] gene number for the *I. scapularis* and *M. occidentalis* genomes were derived from the NCBI-database, while genome size estimates were derived from Ullman *et al.* 2005 and Hoy 2009, respectively

[d] genome size estimate was derived from Ullman *et al*. 2005

[e] these species have been nominated for sequencing by the i5K consortium (http://www.arthropodgenomes.org/wiki/Main_Page)

[f] genome size estimate was derived from Mounsey *et al.* 2012

[g] genome size estimates were derived from Geraci *et al.* 2007

[h] draft genome has been obtained (http://www.uva.nl/over-de-uva/organisatie/medewerkers/content/k/a/m.kant/m.kant.html, status February 15 2013)

2.3.1.1 *Tetranychus urticae* nuclear genome

The genome of the two-spotted spider mite, *T. urticae* (Koch) was recently sequenced and presented as a key-genome for several reasons (Grbić *et al.* 2011). First, spider mites belong to the subphylum of the Chelicerata that comprise the second largest group of animals on this planet next to insects (see 1.1. Introduction). No complete chelicerate genome has been available until recently. Secondly, spider mites can feed on more than 1,100 plant species (including plants toxic to other species) from more than 140 plant families, which is a rare trait amongst arthropods (Bernays and Graham 1988; Migeon and Dorkeld 2010). In addition, they have a high tendency to develop resistance to insect/acaricides of many different mode of action-classes. These features make the two spotted spider mite a very important agricultural pest (see 1.2 Acari of Agricultural importance) (Van Leeuwen *et al.* 2010a).

Noteworthy, the genome of *T. urticae* is 90 Mb long and spread over three chromosomes (Helle and Bolland 1967, Hanrahan and Johnston 2011, Grbić *et al.* 2011). To date, this is the smallest arthropod genome sequenced so far (Table I.2 and Table I.5). Multiple characteristics of the *T. urticae* genome correlate with its compact size: small transposable element content and microsatellite density, increased gene density and holocentric chromosomes. Transposable elements totaled only 9.09 Mb while microsatellite density in the *T. urticae* genome is among the lowest (< 200 microsatellites/Mb) observed for arthropods, consistent with the expectation that repeat content of genomes typically scales with genome size. Furthermore, the gene density of the genome is twice as high compared to *D. melanogaster* with 205 versus 92 genes per Mb. Finally, the holocentric nature (the entire length of the chromosome acts as the centromere) of *T. urticae* chromosomes is correlated with a lack of large tracts of gene-poor heterochromatin. Interestingly, the uniformly distributed gene density contrasts with another arthropod with a small genome, the human body louse *P. humanus*, where 95% of the genes are concentrated in only 55 Mb of the 110 Mb genome (Grbić *et al.* 2011).

2.3.1 Acari transcriptomes

Till now, more than 30 tick transcriptome studies have been published (for a review see Nene *et al*. 2009 and Karim and Adamson 2012). On the other hand, transcriptome data of mites belonging to either the Acariformes or the parasitiform order Mesostigmata are relatively scarce or non-existent (see also Table I.3). Two years ago, two transcriptome studies about

the predatory mite *Phytoseiulus persimilis*, were published (Cabrera *et al.* 2011; Huang 2011). Recently, Roche/454 sequencing was used to sequence the transcriptome of the closely related *Metaseiulus occidentalis* (Hoy *et al.* 2013). Ozawa *et al.* (2012) analyzed transcriptomes of *P. persimilis* and its prey *T. urticae*. Interestingly, the contigs found in this study supported 90.7% (13 257) of the predicted genes and 75.7% (11 062) of the ESTs, reported by the *T. urticae* genome study (Grbić *et al.* 2011). Within the Acariformes, EST-datasets are available for several species of the Astigmata suborder (*Acarus siro*, *Aleuroglyphus ovatus*, *Blomia tropicalis*, *Dermatophagoides* sp., *Glycyphagus domesticus*, *Psoroptes ovis*, *Sarcoptes scabiei, Suidasia medanensis* and *Tyrophagus putrescentiae,* having 18,130 ESTs in total (Table I.3) (Kenyon *et al*. 2003; Ljungren *et al*. 2003; Angus *et al*. 2004; Burgess *et al*. 2011)) while in the suborder of the Prostigmata, three transcriptome studies of different members of the Tetranychidae family have recently been published (Grbić *et al*. 2011; Liu *et al*. 2011a; Niu *et al*. 2012).

3 From Acari genomes to Acari biology

3.1 Phylogenetics

3.1.1 Mitochondrial genomes

Mitochondrial genome sequences have been numerously used to infer phylogenetic relationships (Rubinoff and Holland 2005; Talavera and Vila 2011). In most studies sequence information is used from single mt protein coding genes (PCG). One of these mt PCG, cytochrome oxidase 1 gene (*cox1*) is well known as the choice of locus for DNA barcoding in animals (www.barcodeoflife.org) and has also frequently been used in Acari phylogenetic studies (Navajas *et al*. 1998; Murrell *et al*. 2001a; Walton *et al*. 2004; Dabert 2006; Dabert *et al*. 2010; Arabi *et al*. 2012; Knee *et al*. 2012; Matsuda *et al*. 2012). Although to a lesser extent, other mt sequences that have been used to infer Acari relationships are 12S and 16S ribosomal RNA (Murrell *et al*. 2001ba; Skerratt *et al*. 2002; Walton *et al*. 2004; Humair *et al*. 2007; Chen *et al*. 2012; Nava *et al*. 2012).

Recently, concatenated mt PCG have also been employed to infer phylogenetic relationships within the Acari (see Chapter II and III) (Fahrein *et al.* 2007; Jeyaprakash and Hoy 2009b; Burger *et al.* 2012; Mans *et al.* 2012; Ovchinnikov and Masta 2012). For example, a maximum likelihood phylogenetic analysis using concatenated mt PCG succeeded in recovering Acari relationships concordant with traditional views of Acari phylogeny (see Chapter II). Interestingly, a similar approach has also been used by Rota-Stabelli *et al*. (2010) to infer arthropod relationships at higher taxonomic levels. In this study the monophyletic origin of the Ecdysozoa, a group of molting organisms, comprising the phyla of the Arthropoda, Tardigrada, Onychophora, Nematoda, Nematomorpha, Priapulida, Kinorhyncha and Loricifera (Aguinaldo *et al.* 1997), was highly supported. Nevertheless, other authors believe that the phylogenetic signal contained in mt genomes is limited and should not be used for inferring deep-level arthropod relationships (Curole and Kocher 1999; Cameron *et al.* 2004; Rubinoff and Holland 2005; Talavera and Vila 2011).

Other characters from mt genomes, like gene rearrangements (Boore *et al.* 1998) or inferred secondary structures of rRNA and tRNA molecules (Masta and Boore 2008; Masta 2010) have also been employed to construct phylogenetic relationships. In the case of the Ixodidae family, it was shown that the division of Prostriata and Metastriata could be linked to differences in mt gene arrangement (Black and Roehrdanz 1998; Shao *et al*. 2004). However,

for deduction of phylogenetic relationships between higher Acari (and by extension Arthropod) taxa, mt gene orders seem less useful (see Chapter II) (Cameron *et al.* 2006; Li *et al.* 2012).

3.1.2 Nuclear genomes

In the past, arthropod phylogenetic studies based on nuclear sequence data could only use single (mostly partial) nuclear sequences to infer phylogenetic relationships (Turbeville *et al.* 1991; Friedrich and Tautz 1995; Kjer 2004). In many of these studies, 18S rRNA, 28S rRNA or internal transcribed spacers (ITS) were the sequence of choice. Logically, these sequences have also frequently been used to reveal Acari relationships (Navajas *et al*. 1998; Murrell *et al*. 2001b; Webster *et al*. 2004; Klompen *et al*. 2007; Klimov and Oconnor 2008; Carew *et al*. 2009; Dabert *et al*. 2010; Dowling and Oconnor 2010; Pepato *et al*. 2010; Matsuda *et al*. 2012; Pachl *et al*. 2012; Klimov and OConnor 2013). Although to a lesser extent, EF1-α and heat shock proteins (hsp) have also been used in phylogenetic studies regarding Acari (Domes *et al.* 2007; Klimov and Oconnor 2008; Klimov and OConnor 2013). Remarkably, using a phylogenetic analysis based on three nuclear sequences (18S rRNA, EF1-α and hsp82), Domes *et al.* (2007) could show that an oribatid mite family re-evolved sexuality from a parthenogenetic ancestor. The latter finding violates Dollo's law stating that complex ancestral states can never be reacquired. Another infringement of Dollo's law was recently published by Klimov and OConnor (2013). Among other techniques, these authors used a phylogenetic analysis of five concatenated nuclear genes of more than 300 taxa to demonstrate that house dust mites re-evolved to a free-living state from their permanent parasitic ancestor.

In the last decade, due to the rapid acquisition of large amounts of genomic data using high-throughput genomics, phylogenetics entered a new era: phylogenomics. Whereas conventional phylogenetics are based upon the analysis of a few/single gene(s), phylogenomics investigate evolutionary relationships based on genome-scale data (Eisen and Fraser 2003; Desalle and Rosenfeld 2012; Trautwein *et al.* 2012). In the beginning of this phylogenomic era, all phylogenomic studies of Arthropoda reflected a common sampling tradeoff, using either many genes and very few taxa or too few genes and many taxa. Both approaches suffer from significant sampling biases that can affect inference. It is only recently that arthropod relationships have been studied with a true phylogenomic paradigm spirit (Wheat and Wahlberg 2013). In 2010, for example, Regier *et al*. used 41 kb of aligned

DNA sequence from 62 single-copy nuclear PCG from 75 arthropod species to resolve deep-level arthropod relationships while Meusemann *et al.* conducted a phylogenetic analysis for the same purpose using 129 genes from 117 arthropod taxa. The most integrative approach for deducting arthropod relationships has probably been published in 2013 by Rota-Stabelli *et al*.. These authors present a timescale of ecdysozoan evolution based on five molecular data sets (containing in total 300 unique genes with an average number of taxa of 64) and the most complete set of fossil calibrations. Remarkably, their phylogenetic analyses (again) failed to recover the monophyletic origin of Acari. However, the latter has been questioned for more than 60 years (Dunlop and Alberti 2008; Pepato *et al.* 2010) and only few mite taxa were used in the study of Rota-Stabelli *et al.* (2013). Hence, a thorough phylogenomic analysis focusing on Acari and their closest relatives, with wider sampling of taxa, should be performed to shed more light on this long-debated phylogenetic issue.

3.2 Comparative genomics

As the first completely sequenced and annotated chelicerate nuclear genome, the *T. urticae* genome expands the set of arthropod nuclear genomes beyond crustaceans and insects and provides an important outgroup for comparative genomics. Comparison of the coding gene repertoire of *T. urticae* with the arthropods *T. castaneum*, *D. melanogaster, Nasonia vitripennis* and *D. pulex*, the chordate *H. sapiens* and the cnidarians *Nematostella vectensis* revealed that 4,416 gene families were found to be unique to *T. urticae* and more than 1000 gene families, still present in other arthropods, were lost in *T. urticae*. In addition, many gene families (>50) were significantly expanded compared to other arthropods. In line with its polyphagous nature and the unmatched ability of *T. urticae* to develop resistance, known families implicated in digestion (cysteine peptidases), detoxification (carboxyl/choline esterases (CCEs), glutathione S- transferases (GSTs), cytochrome P450 mono-oxygenases (CYPs)) and transport of xenobiotics (ATP-binding cassette transporters) were among these expanded families. Eighty-six CYP genes were detected in the *T. urticae* genome, a total number similar to insects but with an expansion of *T. urticae* specific intronless genes of the CYP2 clan. The CCE gene family contained 71 genes, with two new clades at the root of the neurodevelopmental class of CCEs representing 34 and 22 CCEs, respectively. Next, a significant expansion was found within the family of 32 GST genes that include a group of mu-class GSTs that were, until now, believed to be vertebrate specific. Finally, 39 multidrug resistance proteins belonging to ABC-proteins (class C) were identified. This number of ABC-C proteins far exceeds those found in other arthropods, vertebrates and nematodes.

Surprisingly, a number of clear cases of horizontal transfer also seems to have drastically expanded the enzymatic potential of spider mites, including a proliferated family of intradiol ring-cleavage dioxygenases (ID-RCDs) that was previously unreported from metazoan genomes (Grbić *et al.* 2011). ID-RCDs catalyze the oxygenolytic fission of catecholic substances, and exclusively allow bacteria and fungi to degrade aromatic rings, a crucial step in the global carbon cycle (Harwood and Parales 1996; Vaillancourt *et al.* 2006). In *T. urticae*, these dioxygenases might have evolved to metabolize aromatic compounds found in plant allelochemicals (Grbić *et al.* 2011). Besides ID-RCDs, other clear instances of lateral gene transfer include (1) the presence of a cobalamin-independent methionine synthase (*MEtE*), (2) two levanase-encoding genes, (3) a cyanate-lyase encoding gene that might be involved in feeding on cyanogenic plants and (4) and a set of carotenoid biosynthesis genes. Remarkably, spider mites are, together with aphids, the only animals containing carotenoid biosynthesis genes in their genome (Moran and Jarvik 2010). Carotenoids play an important role in diapause induction in spider mites (Veerman and Helle 1978; Veerman 1985), and the presence of these genes might indicate that spider mites can also synthesize them (Grbić *et al.* 2011).

3.3 Post-genomics

With a deciphered genome sequence, it is possible to expand the scope of biological investigation from studying single genes or proteins to studying all genes or proteins at once in a systematic way. Various new technologies and disciplines have been developed to assist in addressing biological questions in a genome/proteome-wide way: microarrays, RNAi, next generation sequencing (NGS), RNA-seq, state-of-the-art mass-spectrometry, yeast two-hybrid assays and "(post-gen)omics" disciplines like transcriptomics, peptidomics, proteomics and metabolomics (Hieter and Boguski 1997; Boerjan *et al.* 2012; Palli *et al.* 2012). In the subsections below I will briefly discuss the use of some of these technologies in addressing biological questions concerning Acari biology and toxicology.

3.3.1 Transcriptomics

Transcriptomics or gene expression profiling, examines the expression levels of RNA transcripts in a given organism, hereby often using high-throughput technologies like gene expression microarrays and NGS technologies (Boerjan *et al.* 2012). A typical gene expression (GE) microarray consists of ten thousands of probes complementary to the transcripts whose presence is to be investigated. In the most recent arrays, these probes are

designed based on predicted genes of the genome or open reading frame detected in available transcripts. Transcripts are extracted from samples to be investigated, subsequently labeled with fluorescent dyes and hybridized to the probes on the array, providing information on gene relative transcript ratios (Malone and Oliver 2011; Boerjan *et al.* 2012). GE microarrays have already been developed for several Acari species (see Chapter IV) (Saldivar *et al.* 2008; Aljamali *et al.* 2009; Rodriguez-Valle *et al.* 2010; Bissinger *et al.* 2011; Burgess *et al.* 2012; Demaeght *et al.* 2013). Saldivar *et al.* (2008) used a GE microarray to compare acaricide resistant and susceptible tick strains while Rodriguez-Valle *et al*. (2010) demonstrated that *R. microplus* ticks show different transcript expression patterns when they encounter tick resistant and susceptible breeds of cattle. In another study, Bissinger *et al.* (2011) investigated transcriptome differences between synganglia of unfed and part-fed virgin and replete females of the American dog tick *D. variabilis*. Recently, the first whole genome GE microarray for an Acari species, the two-spotted spider mite *T. urticae* was designed. This GE microarray was used to examine expression differences associated with host plant change, as well as expression patterns in acaricide-resistant strains (see Chapter IV).

Next to gene expression microarrays, quantification of transcriptomes can also be achieved using NGS technologies with either RNA-seq or digital gene expression (DGE or Tag-Seq) (Wang *et al.* 2009; Malone and Oliver 2011; Raz *et al.* 2011). In RNA-seq, transcripts present in the starting material are sampled by direct sequencing using NGS technologies. Sequenced transcripts are then mapped back to a reference genome (or transcriptome) and counted to assess the level of gene expression. Compared to microarray analysis RNA-seq yields much more information, including unpredicted novel transcripts and previously unknown alternatively spliced isoforms. It has also better sensitivity to detect genes with very low or high expression levels (Wang *et al.* 2009; Malone and Oliver 2011; Palli *et al.* 2012). DGE uses a similar approach as RNA-seq but, contrary to the latter, sequence reads are obtained only from the 5' or 3' end of each transcript. This method thus excludes detection of splice and other sequence variants but would have greater sensitivity than RNA-seq (Raz *et al.* 2011).

Expression profiling using NGS technologies is still in its initial phase for mites (Grbić *et al*. 2011; Liu *et al*. 2011a; Niu *et al*. 2012; Ozawa *et al*. 2012). In a more than 3000-fold hexythiazox-resistant *Panonychus citri* strain, laboratory-selected out of a susceptible strain, an RNA-seq analysis revealed that 2,701 genes were differentially expressed (FDR < 0.001,

$|\log_2(FC)| > 1$) compared to the susceptible strain, with heme/copper-type cytochrome/quinol oxidases among the highest upregulated genes (Liu *et al.* 2011a). In 2012, Ozawa *et al.* used pyrosequencing to detect transcriptome differences in both predatory mites and spider mites grown at different temperatures (25°C and 35°C). Not surprisingly, at 35°C genes coding for heat shock proteins were among the most highly differentially expressed genes in both mites. Finally, RNA-seq profiling was used by Grbic *et al.* (2011) to study gene expression patterns in spider mite larvae developing for 12 hours on bean (*Phaseolus vulgaris*), their preferred host, compared to larvae that had been deposited on a more toxic host (tomato (*Solanum lycopersicum*) and *Arabidopsis thaliana*). Genes in the detoxification and peptidase families showed the most profound changes, with expression of nearly half of P450 genes affected by the host plant, including 19 of the 39 genes in the intronless CYP392 and the CYP389 family. A similar pattern was also observed in other families. For GSTs and CCEs the expression of mu and delta GST and the two spider mite specific CCE clades were affected (see 3.2 Comparative genomics). Finally, many genes lacking homology to genes with known function, like putative secreted proteins and lipocalins, were among those genes with the most extreme expression changes upon host transfer.

3.3.2 Proteomics

Proteomics refers to the study of all proteins present in a cell, tissue, organ or organism, including their quantification, identification, localization and modification. In proteomics the mass spectrometry technology has become the method of choice for analysis of complex protein samples ((Aebersold and Mann 2003), for a review see Yates *et al.* 2009). Proteomic technologies have already been frequently used in insect biology, e.g. in the study of insect-plant interactions and in the identification of saliva proteins of major arthropod vectors of human pathogens (Chen *et al.* 2005; Pauchet *et al.* 2008; Carolan *et al.* 2009; Cilia *et al.* 2011; Palli *et al.* 2012; Patramool *et al.* 2012). Within the Acari, several proteomic studies have been performed, with the majority dealing with tick salivary gland proteomes, also called sialomes (from the Greek word *sialo* meaning saliva). According to these studies, tick saliva contains a rich source of proteins and low molecular weight effectors with potent pharmacologic action that may target different mechanisms of coagulation, platelet aggregation and vasoconstriction (Chmelar *et al.* 2012; Patramool *et al.* 2012). In contrast to ticks, mite proteomics are non-existent, which probably is related to their small body (and salivary gland) size. Richards *et al.* (2011a) have probably made the closest attempt to mite proteomics up till now. Electrophoretic analysis of *V. destructor* salivary secretions by SDS-

PAGE and electroblotting indicated the presence of at least 15 distinct protein bands, ranging from less than 17 kDa to 130 kDa. These authors suggested that these protein containing salivary secretions might facilitate the ability of *V. destructor* to feed repeatedly on their bee hosts by suppressing haemocyte-mediated wound healing and inhibiting responses in the host. Similarly to bee-parasite interactions, proteomics could also be a valuable tool to study plant-mite interactions. Spider mites and many eriophyid mites inject saliva into the plant, thereby interfering with the plant hormone equilibrium (Storms 1971; De Lillo and Monfreda 2004; Sarmento *et al.* 2011). Till now proteomics have only been applied on the plant-side of plant-mite interactions (Maserti *et al.* 2011). In addition, it is not known yet whether the saliva of these mites even contains protein substances.

3.3.3 RNA interference

RNA interference (RNAi) or post-transcriptional silencing occurs in a wide variety of eukaryotic organisms, including plants, fungi, arthropods, mammals and nematodes. It is a cellular process by which an mRNA is targeted for degradation by a dsRNA with a strand complementary to a fragment of such mRNA. It has become an important tool for functional genomics in insects (Meister and Tuschl 2004; Belles 2010; Lilley *et al.* 2012). Within the Acari, tick RNAi is the best documented. The majority of proteins involved in the RNAi pathway have been identified in both the *R. microplus* and *I. scapularis* genome/transcriptome and inhibition of protein secretion, decrease in transcript level and reduction in encoded proteins have been demonstrated *in vivo* and *in vitro* for ticks using dsRNA mediated interference. dsRNA for RNAi in ticks can either be administered by injections, soaking, capillary feeding or electroporation (de la Fuente *et al*. 2007a; Kurscheid *et al*. 2009; Karim *et al*. 2010; Karim and Adamson 2012). Contrary to ticks, RNAi research in phytophagous mites is still in its initial phase. A genome-wide screen in the spider mite *T. urticae* revealed that all RNAi machinery (Dicer, Argonautes, RISC components) is present in this mite (Grbić *et al.* 2011). Similar to *I. scapularis*, *T. urticae* also has several duplicated copies of an RNA-directed RNA polymerase (RdRP), an enzyme not yet described in insects (Zong *et al.* 2009; Grbić *et al.* 2011). In nematodes such as *C. elegans*, feeding and soaking in dsRNA results in strong RNAi effects, and RdRP plays a crucial role in the observed systemic effects (Sijen *et al.* 2001). Spider mite RNAi has also been experimentally validated *in vivo*. Injection of dsRNA/siRNA into the spider mite female abdomen was distributed throughout the abdominal cavity and was passed to the egg. Developing embryos inside these eggs displayed the expected RNAi phenotype suggesting strong RNAi mechanisms (Khila

and Grbic 2007). Recently, spider mite RNAi has also been demonstrated through dsRNA delivery via leaf disc feeding. *T. urticae* ingested target dsRNA along with host plant sap and tissues without any antifeedant effect, resulting in a significant knockdown of candidate target-genes. RNAi-mediated knockdown of one of the targeted genes (T-COPB2, Coatomer protein complex, subunit beta 2) even caused more than 60% mortality at 120h post treatment (Kwon *et al.* 2013). Next to spider mites, RNAi has also been shown in the honeybee mite *V. destructor*. RNAi knockdown of a glutathione-S transferase (*VdGST-mu1*) was achieved either by dsRNA injection or overnight immersion of *V. destructor* in dsRNA. Remarkably, Garbian *et al.* (2012) recently demonstrated that RNAi silencing in *V. destructor* could also be achieved by bees ingesting dsRNAs of *V. destructor* gene sequences and transmitting these to the parasitic mite.

3.3.4 DNA methylation

Epigenetic information is an important, environmentally responsive mediator of the relationship between genotype and phenotype, which results from mechanisms other than changes in DNA sequence. Next to histone modification, one of the best-known forms of epigenetic information is DNA-methylation. The key players of the DNA-methylation machinery consist of a family of DNA methyltransferase (dnmt) enzymes, covalently adding a methyl group to cytosine bases. This dnmt family can be divided into three types of enzymes, dnmt1s, dnmt2s and dnmt3s. Dnmt1s or *maintenance* methyltransferases are responsible for maintaining the pattern of DNA methylation through DNA replication while dnmt3s or *de novo* methyltransferases are required for establishing new methylation patterns within in an organism's genome. The function of dnmt2 on the other hand is not well-known. DNA methylation can either occur in transposons, gene bodies (intron/exon) or promoter regions and has been implicated in developmental regulation and many diseases in humans (Glastad *et al.* 2011; Lyko and Maleszka 2011; Jones 2012; Smith and Meissner 2013).

With a few exceptions (Coleoptera, Diptera), DNA methylation has also been demonstrated in most insect lineages but its function is not well understood. In insects, mainly gene bodies and ubiquitously expressed genes are methylated, and it has been suggested that DNA methylation plays an important role in embryogenesis and phenotypic plasticity (Glastad *et al.* 2011; Lyko and Maleszka 2011). The latter was illustrated for the honeybee *Apis mellifera* in which knockdown of a *de novo* methyltransferase gene resulted in the majority of lab-reared larvae to develop a queen phenotype (Kucharski *et al.* 2008) while Zwier *et al.* (2012)

showed that lowering of maternal *dnmt1a* mRNA results in embryonic lethality during the onset of gastrulation in *N. vitripennis*.

The most widely used technique for discovering DNA methylation patterns is bisulfite sequencing, which is based on the property of bisulfite to induce the conversion of unmethylated cytosines to uracil. Hence, sequencing of bisulfite treated DNA will reveal the position of methylated cytosines. Whole-genome bisulfite sequencing is the ultimate DNA methylation analysis technique as it provides single base pair resolution at genome wide coverage but is, however, still too expensive to be applied on large genomes (Morozova and Marra 2008; Laird 2010; Gu *et al.* 2011).

Remarkably, the majority of the DNA methylation machinery is also present in the spider mite *T. urticae*. Both dntm1, 2 and 3 were detected in its genome. The maintenance methyltransferase (dnmt1), however, lacks RNA-seq support and contains a transposable element in its coding region indicating it is probably not functional in *T. urticae*. Nevertheless, DNA-methylation, albeit at low level, has been experimentally validated in the two-spotter spider mite (Grbić *et al.* 2011). In other sequenced mite and tick genomes (see Table I.5) it is not known whether the DNA methylation toolkit occurs. Phenotypic plasticity, the capacity of a single genotype to exhibit variable phenotypes in different environments (Whitman and Agrawal 2009), has also been reported in spider mites (Ito and Saito 2006; Magalhaes *et al.* 2011). A well-known example is the ability of spider mites to go into diapause under unfavorable conditions like shorter photoperiod, lower temperatures and declining food supply (Veerman 1985). As the genome of the spider mite is among the smallest sequenced arthropod genomes, the whole genome bisulfite sequencing approach would be an ideal technique to investigate whether DNA methylation is one of the underlying mechanisms of this biological process.

4 Applied Acari genomics

4.1 Acari and pest management

4.1.1 Resistance research

The intensive use of insecticides and acaricides has led to resistance in many insect and mite species around the world (Table I.6). Important crop pests, parasites of livestock, common urban pests and disease vectors have in some cases developed resistance to such extent that their control not only becomes exceedingly challenging (Van Leeuwen *et al.* 2010a) but also very expensive (e.g. use of higher dosages, substitute with new products). Only in the United States crop losses to pesticide resistance are estimated to be about 1.5 billion per year (Pimentel 2009).

Table I.6 - Top 10 resistant arthropods, based on the number of unique active ingredients for which resistance has been reported and the number of cases reported (Whalon *et al.* 2013)

Species	Taxonomy	Pest type	No. of active ingredients	Cases
Tetranychus urticae	Acari: Tetranychidae	Crop	92	389
Plutella xylostella	Lepidoptera: Plutellidae	Crop	91	544
Myzus persicae	Homoptera: Aphididae	Crop	74	494
Musca domestica	Diptera: Muscidae	Urban	55	279
Leptinotarsa decemlineata	Coleoptera: Chrysomelidae	Crop	54	212
Bemisia tabaci	Homoptera: Aleyrodidae	Crop	49	392
Helicoverpa armigera	Lepidoptera: Noctuidae	Crop	48	671
Panonychus ulmi	Acari: Tetranychidae	Crop	46	189
Boophilus microplus	Acari: Ixodidae	Lifestock	44	167
Blattella germanica	Dermaptera: Blattellidae	Urban	43	219

Today the major emphasis in resistance research lies on unraveling the underlying molecular mechanisms, in an attempt to use this knowledge to control the development and spread of resistant populations (Van Leeuwen *et al.* 2010a). The development of insecticide resistance is influenced by many factors, including genetics, biology/ecology and control operations (Georghiou and Taylor 1977). There are many possible adaptations that permit an insect or mite to survive lethal doses of an insecticide/acaricide. These are usually classified based on their biochemical/physiological properties, as either mechanisms of decreased response to the pesticides (interaction of a pesticide with its target site), or mechanisms of decreased exposure (penetration, distribution, metabolism and excretion). The majority of cases involve

changes in the sensitivity of the target site due to point mutations, or sequestration/metabolism of the insecticide before it reaches the target site due to quantitative or qualitative changes in major detoxification enzymes (CCEs, CYPS and GSTs) (Van Leeuwen *et al.* 2010a).

In the past, the identification of putative target-sites and resistance genes has been limited to those cases where genes could be identified by degenerate PCR, RACE and cloning. The availability of a genome sequence is of now great help in identifying targets and resistance mechanisms, first by simply cataloguing the relevant genes and gene families, as has recently been done for the spider mite *T. urticae* (see 3.2 Comparative genomics, Chapter V and VI) (Grbić *et al.* 2011). Next, comparative expression profiling of sensitive and resistant strains could identify candidate genes associated with resistance, without presumption of the resistance mechanisms. Daborn *et al*. first conducted this approach in 2002 for insecticide resistance in *D. melanogaster*. Recently, similar approaches have been performed for mites and ticks (see Chapter IV, see 3.3.1 Transcriptomics). For example, using microarray technology, Demaeght *et al.* (2013) demonstrated that a member of the CYP2 clan was associated with spirodiclofen resistance in *T. urticae* while Saldivar *et al.* (2008) showed that a glutathione-S transferase was differentially upregulated in the cattle tick *R. microplus* upon exposure to coumaphos.

In establishing molecular mechanisms of mite and tick resistance, the lack of genetic resources has also been a limiting factor for the direct cloning of genes underlying resistance. The potential of such resources was recently illustrated for the spider mite *T. urticae*. Analysis of the full mt genome sequence has lead to the documentation of the mode of action of bifenazate (Van Leeuwen *et al.* 2008) while in 2012 the complete nuclear genome allowed the relatively fast discovery of a new mutation associated with abamectin resistance in *T. urticae* (see Chapter VI). Furthermore, the lack of a nuclear genome sequence together with the small size of most Acari have made traditional bi-parental genetic mapping difficult. For example, a preliminary linkage map has only been published for a tick*, Ixodes scapularis* (Ullmann *et al.* 2003). Now that Acari genomic resources are available, genetic mapping should nonetheless be possible. This was recently demonstrated by Van Leeuwen *et al.* (2012). Using a bulk segregant analysis (Michelmore *et al.* 1991) combined with high-throughput sequencing, they identified a locus for monogenic, recessive resistance to etoxazole in the spider mite *T. urticae*. This high-resolution mapping is probably the most

efficient way to identify resistance mechanisms without any pre-assumptions, on condition that the genetic mechanisms underlying resistance are relatively small. However, due to the cost of sequencing, such an approach is currently only affordable with small Acari genomes. In addition, a short generation time and high fecundity of the Acari species under investigation is also a prerequisite to render this BSA approach practically feasible.

4.1.2 Exploiting mite-plant and tick-host-pathogen interactions

A main obstacle for the study of plant-pest interactions is the lack of an easy-to-rear generalist herbivore model organism, which preferentially feeds on a range of established model plants, and for which extensive genomic resources have been developed. The available genome sequence of *T. urticae*, together with the fact that these mites feed on model plants such as *Arabidopsis*, *Medicago* and tomato, offers unique possibilities to investigate mite-plant interactions at both trophic levels. Interestingly, spider mites like *T. urticae* seem to actively manipulate plant defenses, but the molecular mechanisms by which this is accomplished are still unknown (Storms 1971; Kant *et al.* 2008; Sarmento *et al.* 2011). A better understanding of these mechanisms could be exploited to (re)introduce or improve the plants' own natural abilities to control pests (Bleeker *et al.* 2012).

Contrary to mite-plant interactions, tick-host-pathogen relationship have already been extensively studied the last decade. Based on the acquired knowledge several tick antigens with the potential to serve as vaccines to impair feeding and transmission of pathogens have been identified (Willadsen 2004; de la Fuente and Kocan 2006; de la Fuente *et al*. 2007b; Kiss *et al*. 2012a; Radolf *et al*. 2012; Embers and Narasimhan 2013). For example, in Latin-American countries and Australia a commercial vaccine, containing the recombinant *R. microplus* BM86 gut antigen, was registered in the nineties for its use in reduction of *R. microplus* infestations on cattle. However, one of the major drawbacks of this vaccine was the lack of cross-protection against other economically important tick species (e.g. *Amblyomma* sp. and *Dermacentor* sp.), indicating the need for discovery of more conserved tick-protective antigens that could be used to control infestations by multiple tick species (de la Fuente *et al*. 2007b; Guerrero *et al*. 2012; Kiss *et al*. 2012a). In this light, complete nuclear genomes of three *Amblyomma* sp. and *Dermacentor variabilis*, that have recently been proposed for sequencing (see 2.3.1 Acari nuclear genomes), could be a major aid in identifying protective homologues in a range of tick species for maximizing vaccine efficiency against all tick species targeted.

4.1.3 RNAi to control Acari pests?

The potential of RNAi to selectively protect plants from herbivory has been recognized for many years, but a proof of principle example of insect resistance in genetically modified agricultural crops has been presented only fairly recently (Baum *et al.* 2007; Price and Gatehouse 2008; Mao *et al.* 2011; Pitino *et al.* 2011). As RNAi silencing in spider mites has recently been achieved by administering dsRNA through leaf disk feeding, plant mediated mite RNAi could potentially also be a novel tool for controlling spider mites. For this purpose, previously constructed RNA-seq libraries of *T. urticae* (Grbić *et al.* 2011) could be screened for optimal RNAi targets and current available genomic resources of spider mite predators like *P. persimilis* and *M. occidentalis* allow to select only those target genes that are unique to *T. urticae*.

RNAi also proved to be a promising concept for control of the honeybee mite *V. destructor*. Recently, Garbian *et al.* (2012) demonstrated that bees ingesting dsRNA of *V. destructor* gene sequences become vectors of dsRNAs and transmit these to the honeybee mite, triggering mite gene silencing and resulting in a more than 50% reduction of the mite population. Noteworthy, the dsRNA *V. destructor* gene sequences could be easily administered to the colony in a sugar solution and did not affect the vigor of the bees. However, long-term effects of dsRNA treatment of honey bee colonies and optimization of RNAi mite target genes, leading to a greater reduction of the mite population, should be performed before *V. destructor* RNAi can be used as an effective control method in the field.

Finally, RNAi mediated control of ticks has also been proposed. The subolesin gene is a highly conserved gene across tick species. RNAi silencing of this gene in ticks disrupts reproductive development in both female and male ticks, resulting in production of sterile ticks. When the latter are crossed with normal ticks, lack of successful mating by the sterile ticks prevents female engorgement and oviposition. Consequently, it was suggested that release of subolesin silenced ticks may have applications in reducing populations of many tick species. In analogy with the sterile insect technique (SIT), this method was named the sterile acarine technique (SAT) (de la Fuente *et al*. 2006; Merino *et al*. 2011; Kiss *et al*. 2012a). However, while the SIT-technique has been proven to be successful for controlling many insect crop pests and pathogen vectors (Dyck *et al.* 2005), the potential of SAT in tick control has not yet been explored in large-scale field tests (Kiss *et al*. 2012a).

4.2 Industrial and medical applications of Acari molecules?

Now that complete Acari genomes are available, these genomes can also be mined for genes having potential in industrial and medical applications. The spider mite *T. urticae* for example contains a set of intradiol ring-cleavage dioxygenases (ID-RCDs), an enzyme family that, until recently, was believed to be absent in metazoan species (Grbic *et al.* 2011). ID-RCDs allow bacteria and fungi to degrade aromatic compounds with some ID-RCDs even capable of decomposing man-made aromatic environmental pollutants like PCBs. As such these enzymes might have great potential in bioremediation of pollutants (Fortin *et al.* 2005; Scott *et al.* 2008; Singh *et al.* 2013). Whether spider mite ID-RCDs also have such potential is an open question but it would be worth to be further explored (see Chapter IV). Next to ID-RCDs, *T. urticae* was also reported to possess a set of antifreeze proteins (AFPs) (Bryon *et al.* 2013). In several animals, e.g. fish in polar oceans, these AFPs help to prevent ice formation at subzero temperatures. Based on the latter property, many applications have been suggested for AFPs. Adding AFPs for example to food may inhibit recrystallization during freezing storage, transport, and thawing and hence preserve food texture by reducing cellular damage and minimizing the loss of nutrients by reducing drip (Venketesh *et al.* 2008). Future studies should point out if *T. urticae* AFPs would also be appropriate for this purpose.

Like most other blood feeding animals, tick saliva contains a wide range of physiologically active molecules that are crucial for attachment to the host or for the transmission of pathogens. Some of these active molecules are known to interact with host processes like coagulation and immunity and have as such potential as future anticoagulant (anti-hemostatics) or immunosuppressive agents (Hovius *et al.* 2008). The last decades heparin has long been the anticoagulant of choice for treating thrombosis. Its clinical practice is nonetheless associated with a number of drawbacks and potentially fatal risks. Thus, new anticoagulant agents, like those from ticks, overcoming the limitations of heparin are more than welcome (Laux *et al.* 2009). However, due to the vast number of criteria these compounds must meet before clinical use, tick-based anti-hemostatics have not yet been registered (Maritz-Olivier *et al.* 2007; Chmelar *et al.* 2012).

Chapter II

The complete mitochondrial genome of the house dust mite *Dermatophagoides pteronyssinus* (Trouessart): a novel gene arrangement among arthropods

---

This chapter has been redrafted from:

Dermauw W., Van Leeuwen T., Vanholme B., and Tirry L. (2009). The complete mitochondrial genome of the house dust mite *Dermatophagoides pteronyssinus* (Trouessart), a novel gene arrangement among arthropods. BMC Genomics 10, 107.

1 Introduction

Acari, mites and ticks, diversified 400 million years ago and currently two major lineages are recognised: Acariformes and Parasitiformes (Krantz and Walter 2009). The Acariformes comprise two major groups, the Trombidiformes and Sarcoptiformes (OConnor 1984; Walter et al. 1996). Two of the most prominent members of the Sarcoptiformes are the European house dust mite *Dermatophagoides pteronyssinus* (Trouessart, 1897) and the American house dust mite *Dermatophagoides farinae* (Hughes, 1961), both belonging to the family of the Pyroglyphidae (cohort Astigmata). Pyroglyphid mites are typical inhabitants of animal nests. In the human environment, they are mainly found in upholstery, textile floor covers and beddings, where they primarily feed on the skin scale fraction in house dust (Spieksma 1997). About 40 years ago, house dust mites were first recognized as one of the major sources of allergens in house dust (Voorhorst *et al.* 1967). The allergenic proteins are found in high concentrations in mite faeces, which, after drying and pulverizing, become airborne and can be inhaled. The presence of these allergens in sensitive persons is able to cause diseases like asthma, dermatitis and rhinitis (van Bronswijk 1981; Arlian and Platts-Mills 2001). In countries with a temperate climate, 6 to 35 per cent of the population is sensitive to house dust mite-derived allergens (Janson *et al.* 2001).

Complete mitochondrial (mt) genome sequences are becoming increasingly important for effective evolutionary and population studies. Mt genome sequences are not only more informative than shorter sequences of individual genes, they also provide sets of genome-level characters, such as the relative position of different genes, RNA secondary structures and modes of control of replication and transcription (Dowton *et al.* 2002; Boore *et al.* 2005; Boore 2006a; Masta and Boore 2008). However, the applicability of mt genomes as a marker of highly divergent lineages is still controversial (Curole and Kocher 1999; Cameron *et al.* 2004) and remains to be elucidated (Castro and Dowton 2007). In addition, unravelling mt genomes can be of economic importance as well, since several chemical classes of pesticides target mt proteins. Well-known acaricides like acequinocyl and fluacrypyrim affect mt electron transport through the inhibition of the mt encoded cytochrome b in complex III (Dekeyser 2005). Also, the economically important class of METI (Mitochondrial Electron Transfer Inhibitors)-acaricides target the mt complex I, although their exact molecular target has not yet been elucidated. Recently, resistance to the acaricide bifenazate was shown to be

caused by mutations in the mt encoded cytochrome b and to have evolved rapidly through a short stage of mt heteroplasmy (Van Leeuwen *et al.* 2008).

In 2009, at the time this study was carried out, the mt genomes of 20 species belonging to the Acari were available at NCBI (www.ncbi.nlm.nih.gov, status January 10, 2009). These deposited sequences have the typical features of metazoan mt genomes. They are circular, between 13 and 20kb in length, contain a coding region with 37 genes (22 tRNAs, 2 rRNAs and 13 protein coding genes) and a relatively small non-coding region. The latter is mostly AT-rich and fulfils a role in the initiation of replication and transcription (Wolstenholme 1992; Boore 1999). Compared to this typical configuration, the mt genomes of *Steganacarus magnus*, *Metaseiulus occidentalis* and *Leptotrombidium pallidum* show some abnormal features. *S. magnus* lacks 16 of the 22 tRNAs normally present in mt genomes (Domes *et al.* 2008). *M. occidentalis* has a unusually large mt genome (24.9 kb) resulting from a duplication event of a large fragment of the codon region. Despite its large size, genes coding for nad6 and nad3 were not found during the initial annotation process (Jeyaprakash and Hoy 2007). *L. pallidum* on the other hand has 38 mt genes due to a duplication of the *16S-rRNA* (Shao *et al*. 2005a)

In this chapter, we analyse the complete mt genome of a member of the Sarcoptiformes, the European house dust mite *D. pteronyssinus*, after obtaining the complete sequence using a long PCR approach.

2 Materials and methods

2.1 Mite identification

Upon arrival in the laboratory, mites were identified as *D. pteronyssinus* by J. Witters (ILVO, Belgium) and F. Th. M. Spieksma (Laboratory of Aerobiology, LUMC, The Netherlands) using morphological characteristics. To back up this identification, molecular techniques were applied. For this purpose DNA was extracted and used as a template for PCR. Primers 12SID-F and 12SID-R (see Table II.1 for primer sequences) successfully amplified a 316bp fragment. BLASTn searches against non-redundant nucleotide sequences using the amplified fragment as query resulted in a perfect match with a mt *12S-rRNA* sequence of *D. pteronyssinus* (GenBank accession number AF529911).

2.2 Mite strain, mass rearing and isolation

The initial *D. pteronyssinus* culture was provided by D. Bylemans (Janssen Pharmaceutica, Belgium). Mites were cultured on a 1:1 mixture of Premium Gold (Vitacraft, Germany) and beard shavings at 75% R.H., 25°C and permanent dark conditions (Miyamoto *et al.* 1975; Kalpaklioglu *et al.* 1996). Mites were isolated from the colony using a modified heat-escape technique (Bischoff 1988; Hart 1990). Briefly, mite cultures were transferred to small plastic petri dishes (75 mm in diameter, 28 mm high) with a lid on top. These dishes were placed in the dark on a hot plate set at 45°C (Bekso, Belgium). After 15-20 minutes the mites moved away from the heat source, formed groups on the lid of the petri dish and could be collected using a fine hair brush.

2.3 DNA extraction

Approximately 1000 *D. pteronyssinus* mites were collected in an Eppendorf tube and were ground in 800 µl SDS-lysis buffer (400 mM NaCl, 200 mM TRIS, 10 mM EDTA, 2% SDS) using a small sterile plastic pestle (Eppendorf, Germany). After incubation for 30 min at 60°C under continuous rotation, a standard phenol-chloroform extraction was performed (Sambrook and Russel 1987). Total genomic DNA was precipitated with 0.7 volumes of isopropanol at 4°C for 1 hour, centrifuged for 45 minutes at 21,000 x g and washed with 70% ethanol. Precipitated DNA was resolved in 50 µl 0.1 M Tris pH 8.2.

2.4 PCR

Standard PCR (amplicon < 500bp) was performed in 50 µl volumes (38.5 µl double-distilled water; 5 µl buffer; 2 mM $MgCl_2$; 0.2 mM dNTP-mix; 0.2 µM of each primer; 1 µl template DNA and 0.5 µl *Taq* polymerase (Invitrogen, Belgium). PCR conditions were as follows: 2' 94°C, 35 x (20" 92°C, 30" 53°C, 1' 72°C) and 2' 72°C. The annealing time was extended to 1 minute and the primer concentration was increased to 2 µM when degenerate primers were used. Long PCR (amplicon > 500bp) was performed with the Expand Long Range Kit (Roche, Switzerland) in 50 µl volumes (28.5 µl double-distilled water; 10 µl buffer; 0.5 mM dNTP-mix; 0.3 µM of each primer; 4 µl 100% DMSO; 1 µl template DNA and 1 µl enzyme-mix). PCR conditions were: 2' 94°C, 10 x (10" 92°C, 20" at a temperature that varies depending on the primers, 1'/kb 58°C), 25 x (10" 92°C, 20" at a temperature that varies depending on the primers, 1'/kb 58°C with 20'' added for every consecutive cycle) and 7' 58°C. All PCR products were separated by electrophoresis on a 1% agarose gel and visualised by EtBr staining. Fragments (amplicon < 1000bp) of interest were excised from gel, purified with the QIAquick PCR Purification Kit (Qiagen, Belgium) and cloned into the pGEM-T vector (Promega, Belgium). After heat-shock transformation of *E. coli* (DH5α) cells, plasmid DNA was obtained by miniprep and inserted fragments were sequenced with SP6 and T7-primers. Long PCR products were sequenced by primer-walking. All sequencing reactions were performed by LGS (formerly known as AGOWA) sequencing service.

2.5 Amplification of the mt genome

Primers COXI-F and 12S-R, based on partial *D. pteronyssinus cox1* and *12S-rRNA* sequences (GenBank accession number: AY525570 and AF529911, respectively) (see Table II.1 for primer sequences), successfully amplified a 4.6kb sequence of the mt genome of *D. pteronyssinus*. Degenerate primers CYTB-F-Deg and CYTB-R-Deg (see Table II.1 for primer sequences), designed on conserved regions of Acari *cytB*, amplified a partial *cytB* sequence from *D. pteronyssinus*. A specific primer COXI-R, designed from the 3' end of the 4.5kb sequence in combination with the primer CYTB-F, designed from the partial *cytB* sequence, successfully amplified a 2.2kb sequence. Another primer CYTB-R, designed from the 5'-end of this 2.2kb sequence, in combination with the primer 12S-F successfully amplified a 8.6kb sequence, making the mt genome sequence complete.

Table II.1 - Primers and their sequences used to characterize the *D. pteronyssinus* mt genome

Primer	Sequence (5'-3')	T_m (°C)
12SID-F	TTTTTCTGGCGGTTTATTACC	58
12SID-R	CGACTTATCTATCAAAAGAGTGACC	59
COXI-F	GCTGTTCCTACGGGAGTTAAGG	61
COXI-R	ACACGCCCTCTTCTCATACC	59
12S-F	GCGGTTTATTACCCATTCACAGG	61
12S-R	ACGGGCGATATGTACTTTTATTAGG	63
CYTB-F-Deg	TAWRAARTATCAYTCDGGTTKRATATG	
CYTB-R-Deg	CCWTGAGGACAAATAWSWTTYTGAGG	
CYTB-F	CGATTCAAGGTGGGTTGG	60
CYTB-R	GATATTGACCACGGTTGAATTATGC	62
DP-Ms-F	TGCTAAAGCAAGAGAGAACTTTT	58
Dp-Ms-R	CCCCTTCAAACCACAACACT	60

2.6 Annotation and bioinformatics analysis

The complete genomic sequence was assembled and annotated using VectorNTI (Invitrogen, Belgium) according to Masta and Boore (2004). Open reading frames (ORFs) were identified with the program Getorf from the EMBOSS-package (Rice *et al.* 2000). The obtained ORFs were used as query in BLASTp (Altschul *et al.* 1997) searches against the non-redundant protein database at NCBI. Two large non-protein-coding regions were candidates for the rRNAs (16S and 12S respectively). The boundaries were identified based on alignments and secondary structures of rRNA genes of other mite species. Sixteen of the 22 tRNAs were identified by tRNA-scan SE (Lowe and Eddy 1997) with a cove cutoff score of 0.1 and the tRNA-model set to "nematode mito". The remaining tRNAs (*trnM*, *trnV*, *trnY*, $trnS_1$, *trnI*, *trnC*) were determined in the unannotated regions by sequence similarity to tRNAs of other mite species. In order to obtain additional information on mt gene boundaries, BLASTn (Altschul *et al.* 1997) searches of *D. pteronyssinus* tRNA, rRNA and protein encoding nucleotide sequences were carried out against ESTs (Gissi and Pesole 2003) restricted to *Dermatophagoides* sequences (n=3532). ESTs with statistically significant matches (E-value cutoff: 0.1) were collected, checked for vector contamination and aligned by Clustal W (Thompson *et al.* 1994) as implemented in BioEdit 7.0.1 (Hall 1999) against the appropriate nucleotide sequence of *D. pteronyssinus.* MatGAT 2.02 was used to calculate similarity and identity values (Campanella *et al.* 2003) of mt proteins. The identification of gene subsets that appear consecutively in different genomes was performed by common interval distance

analysis using CREx (Bernt *et al.* 2007) (see Appendix II-D for input data of the CREx program).

2.7 Construction of secondary structures of RNAs and non-coding regions

Secondary structures of tRNAs were determined following the method of Masta and Boore (2004). Secondary structures of tRNAs were drawn with CorelDraw 12.0 (Corel Corporation, Canada). The rRNA genes of *D. pteronyssinus* were aligned with those of other Acariformes and conserved areas were identified. These regions were mapped on the published structures of *L. pallidum* rRNA (Shao *et al.* 2006). Regions lacking significant homology were folded using Mfold (Zuker 2003). Secondary structures of rRNAs were drawn using the RnaViz2 program (De Rijk *et al.* 2003) and afterwards modified with CorelDraw 12.0 (Corel Corporation, Canada). Secondary structures of non-coding regions were folded using Mfold (Zuker 2003). When multiple secondary structures were possible, the most stable (lowest free energy (-ΔG)) one was preferred. Drawing and editing of these structures was done in a similar way as for rRNA secondary structures.

2.8 Multiply-primed rolling circle amplification (RCA) and restriction enzyme digestion

With multiply-primed RCA an exponential amplification of small amounts of circular DNA can be obtained (Dean *et al*. 2001) (Fig. II.1). Extraction and RCA of mitochondrial DNA of

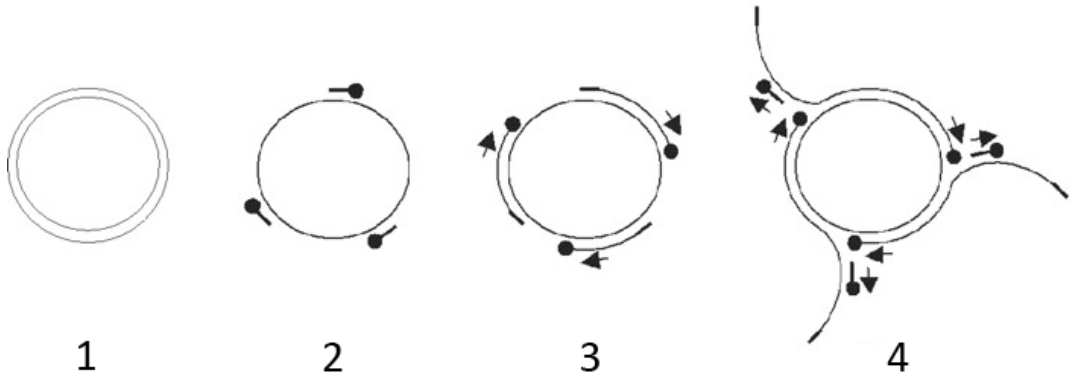


Figure II.1 - Multiply-primed RCA of circular DNA (drafted from Rector *et al.* 2004)

(1) Circular DNA is rendered single stranded by a denaturation step; (2) Exonuclease-protected random hexamer primers (-) anneal to multiple sites on the template circular DNA (3) φ29 DNA polymerase (•) binds and extends these primers at the 3′ end; (4) Strand displacement synthesis occurs when the DNA polymerase reaches a downstream extended primer, and hexamer primers can anneal to the displaced single-stranded product strands and will again be elongated by the φ29 DNA polymerase. Continuation of this process results in exponential amplification of the template DNA, generating linear double-stranded, high-molecular-weight repeated copies.

D. pteronyssinus was done according to Van Leeuwen *et al.* (2008). Multiply-primed rolling circle amplified mtDNA was digested with two enzymes (*Xmn*I and *Eco*RI; New England Biolabs) following the manufacturer's instructions. Restriction digests were fractionated by agarose gel electrophoresis as described before.

2.9 Phylogenetic analysis

Sequence data were obtained from 21 Acari species (for GenBank accession numbers see Table II.2) and two outgroup taxa (*Limulus polyphemus* (GenBank: NC_003057 accession number) and *Locusta migratoria* (Genbank accession number: NC_001712)). Only mite species with a completely sequenced mt genome were selected. Alignments from all mt protein-coding genes were used in phylogenetic analysis. Amino acid sequences and nucleotide sequences were aligned by Clustal W (Thompson *et al.* 1994) as implemented in BioEdit 7.0.1 (Hall 1999). The nucleotide alignment was generated based on the protein alignment using codon alignment. Ambiguously aligned parts were omitted from the analysis by making use of Gblocks 0.91b (Castresana 2000), with default block parameters except for changing "allowed gap positions" to "with half". Abascal *et al.* (2006) recently presented evidence that some insects and ticks use a modified mitochondrial code, with AGG coding for lysine rather than serine as in the standard invertebrate mitochondrial code. As 10 out of 20 Acari species in our dataset are ticks all positions aligning to AGG codons in the final amino acid alignment were removed.

For the nucleotide alignments the "codons" option was used in Gblocks 0.91b (Castresana 2000). Due to the results of a saturation analysis (Xia *et al.* 2003) on single codon positions, implemented in DAMBE 4.2.13 (Xia and Xie 2001), third codon positions were eliminated from the nucleotide alignment. An incongruence length difference test (ILD-test) (Farris *et al.* 1995) as implemented in PAUP* (version 4.0b10; (Swofford 2003)) was used to assess congruence among gene partitions.

Model selection was done with ProtTest 1.4 (Abascal *et al.* 2005) for amino acid sequences and with Modeltest 3.7 (Posada and Crandall 1998) for nucleotide sequences. According to the Akaike information criterion, the mtART+G+I+F model was optimum for phylogenetic analysis with amino acid alignments and the GTR+I+G model was optimal for analysis with nucleotide alignments.

Two different analyses were performed. (1) Maximum likelihood (ML) analysis was performed using Treefinder (Jobb *et al.* 2004), bootstrapping with 1,000 pseudoreplicates and (2) Bayesian inference (BI) was done with MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001).

As the mtART model is not implemented in the current version of MrBayes, the mtREV+G+I model was used for phylogenetic analysis with the amino acid alignment. Four chains ran for 1,000,000 generations, while tree sampling was done every 100 generations. Burnin was calculated when the average standard deviation of split frequencies had declined to < 0.01. The remaining trees were used to calculate Bayesian posterior probabilities (BPP).

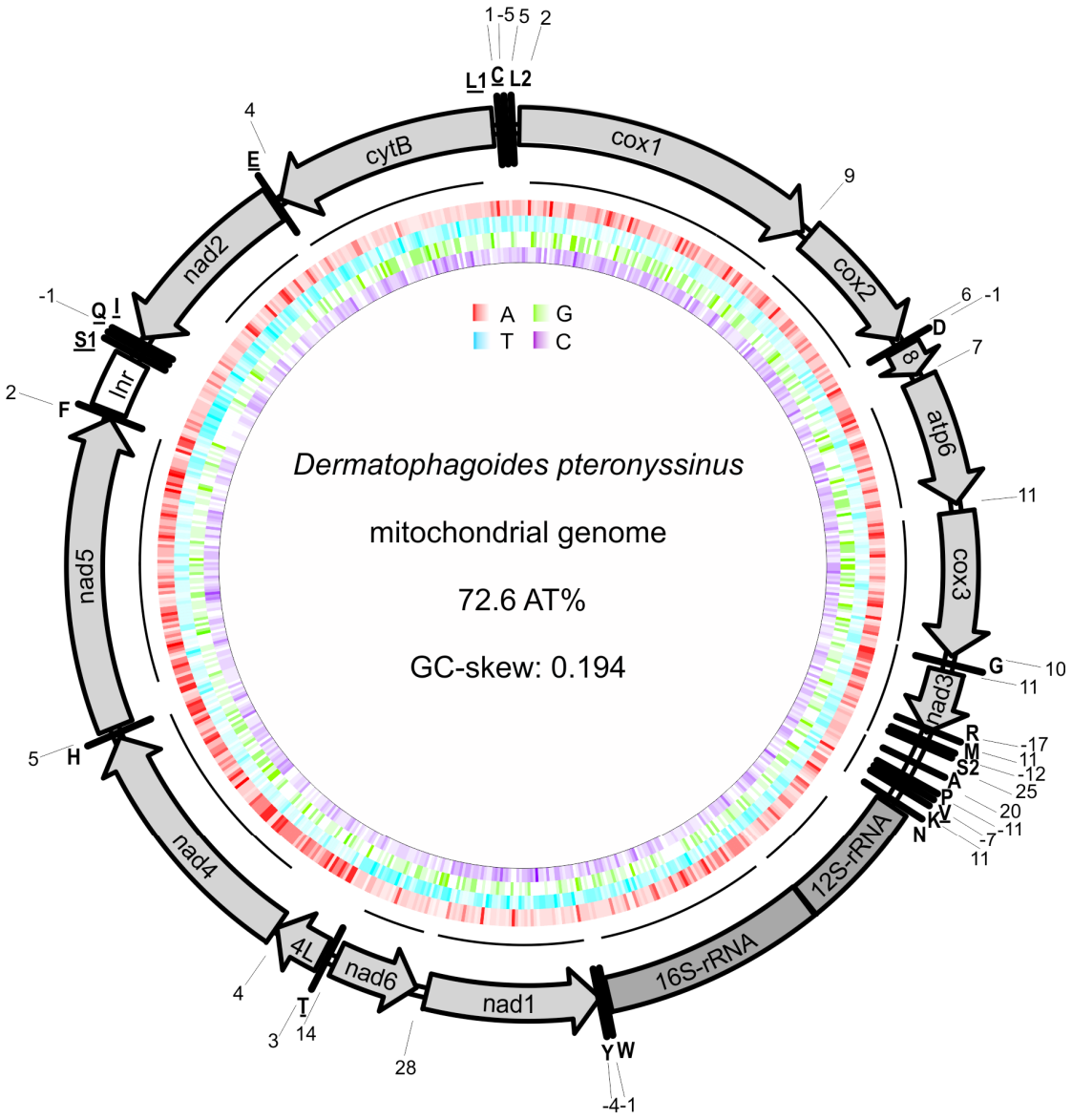


Figure II.2 - Schematic representation of the mt genome of *D. pteronyssinus*.
Except for *atp8* (=8) and *nad4* (=4L) protein coding and ribosomal genes are presented as outlined in the abbreviations section. tRNA genes are abbreviated using the one-letter amino acid code, with L_1=CUN; L_2=UUR; S_1= AGN; S_2=UCN. RNAs on the N-strand are underlined. Numbers at gene junctions indicate the length of small non-coding regions where negative numbers indicate overlap between genes. A-,T-,G- and C-content of the mt genome is represented using a red, blue, green and purple colour graded circle, respectively. Black curved lines on the outside of these circles represent mt genome coverage by *Dermatophagoides* ESTs (see Appendix II-C for sequences of *Dermatophagoides* ESTs covering the mt genome of *D. pteronyssinus*).

3 Results and discussion

3.1 Genome organisation

The mt genome of *D. pteronyssinus* was amplified, using long PCR, in three overlapping fragments and has a final assembled sequence of 14,203bp (GenBank accession number: EU884425; Fig. II.2). As nonspecific amplification artefacts and incomplete coverage of genes are well-known drawbacks of a PCR approach (Gorrochotegui-Escalante and Black 2003), we checked the genome size by restriction digest on rolling circle amplified mtDNA (Fig. II.3). This approach confirmed the sequence size, considering that the relative mobility of mtDNA restriction fragments can show slight (5-12%) deviations compared to their sequence length (Howell 1985). In 2009, the mt genome of *D. pteronyssinus* represented the first mt sequence of a mite belonging to the Astigmata and was, together with the mt genome of *S. magnus,* the only representative from the order of the Sarcoptiformes. Adding this genome to the database resulted in 21 publicly available Acari mtDNA sequences (Table II.2). Twelve belong to species in the superorder of the Parasitiformes whereas nine - among which *D. pteronyssinus* - belong to species in the superorder of the Acariformes.

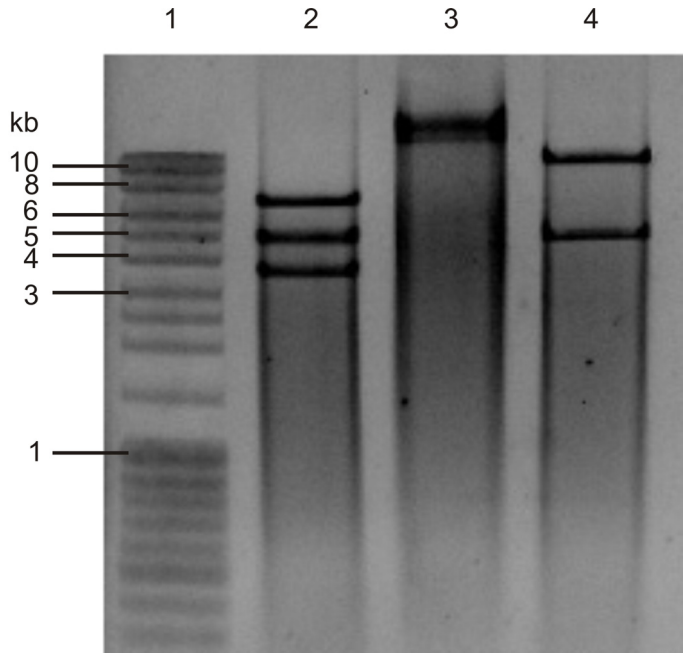


Figure II.3 - Restriction digest of rolling circle amplified mt DNA of *D. pteronyssinus*.
Rolling circle amplified mtDNA, undigested (lane 3) and digested with *Xmn*I (lane2) and *Eco*RI (lane 4). Molecular marker used was MassRuler DNA ladder Mix (Fermentas) (lane 1).

Table II.2 - Nucleotide composition of completely sequenced mt genomes of Acari and *Limulus polyphemus*[*]

Organism [a]	S[b]	Accession nr	Complete mt genome				mt PCG[c]		12S-rRNA		16S-rRNA		Control Region(s)[d]		Ref. [e]
			Length (bp)	AT%	AT-skew[f]	GC-skew[f]	Length bp)	AT%	Length (bp)	AT%	Length (bp)	AT%	Length (bp)	AT%	
D. pteronyssinus	A	EU884425	14,203	72.60	-0.199	0.194	10,826	71.61	665	72.93	1,078	76,07	286	91.61	this study
Am. triguttatum	P	NC_005963	14,740	78.40	-0.022	-0.133	10,876	78.29	693	79.65	1,199	81.82	307-307	71.66-71.01	(un)
As. sp.[g]	A	NC_010596	16,067	70.07	0.015	-0.049	10,560	69.27	680	72.79	1,047	76.31	1,207-1,236	70.01-70.06	(un)
C. capensis	P	NC_005291	14,418	73.54	0.036	-0.374	10,873	72.66	695	76.26	1,225	78.29	342	71.35	(1)
H. flava	P	NC_005292	14,686	76.91	-0.018	-0.116	10,817	76.65	699	78.40	1,196	81.61	310-310	66.45-66.77	(1)
I. hexagonus	P	NC_002010	14,539	72.66	0.033	-0.366	10,826	71.13	705	78.44	1,287	72.60	359	71.87	(2)
I. holocyclus	P	NC_005293	15,007	77.38	-0.013	-0.254	10,860	76.38	716	77.93	1,214	81.55	352-450	78.41-80.00	(3)
I. persulcatus	P	NC_004370	14,539	77.34	-0.024	-0.269	10,879	76.59	720	78.89	1,206	79.77	352	77.56	(3)
I. uriae	P	NC_006078	15,053	74.79	0.007	-0.328	10,837	73.75	712	78.09	1,210	78.35	388-476	77.06-74.16	(3)
Le. akamushi	A	NC_007601	13,698	67.47	-0.016	-0.075	10,292	67.19	596	67.11	1,026	72.03	260-262	60.38-59.54	(4)
Le. deliense	A	NC_007600	13,731	69.95	-0.017	-0.058	10,292	70.06	602	70.27	1,023	73.02	294-301	62.24-61.79	(4)
Le. pallidum[h]	A	NC_007177	16,779	70.96	-0.031	-0.044	10,312	71.38	601	72.05	1,008	74.90	537-724-736-803	63.87-66.71-66.75-66.50	(5)
M. occidentalis[i]	P	NC_009093	24,961	75.97	0.095	-0.291	10,014	74.38	742	81.13	1,192	84.31	310-311-311-311	79.35-79.10-79.42-78.78	(6)
O. moubata	P	NC_004357	14,398	72.26	0.067	-0.379	10,890	71.35	686	74.20	1,212	76.90	342	71.64	(7)
O. porcinus	P	NC_005820	14,378	70.98	0.059	-0.355	10,877	70.11	691	74.38	1,207	74.48	338	69.53	(8)
R. sanguineus	P	NC_002074	14,710	77.96	-0.034	-0.098	10,803	77.96	687	79.18	1,190	81.34	303-305	67.33-66.56	(2)
S. magnus	A	NC_011574	13,818	74.59	-0.020	-0.037	10,560	74.44	609	74.38	992	74.38	1018	75.66	(9)
T. urticae	A	NC_010526	13,103	84.27	0.026	-0.016	10,226	84.00	646	85.91	991	85.27	44	95.45	(10)
U. foili[g]	A	NC_011036	14,738	72.95	0.201	-0.279	10,679	71.83	649	74.35	1,016	74.35	387-644	76.49-77.33	(un)
V. destructor	P	NC_004454	16,477	80.02	-0.021	0.177	10,728	79.22	726	80.44	1,149	83.12	2174	79.71	(11)
W. hayashii[g]	A	NC_010595	14,857	72.97	0.264	-0.305	10,573	73.01	625	75.05	1,045	77.42	1403	68.28	(un)
Li. polyphemus	[j]	NC_003057	14,985	67.60	0.111	-0.399	11,077	66.43	799	69.70	1,296	71.00	348	81.3	(12)

[*] values were obtained from the corresponding GenBank flat-file in the NCBI database (status January 10, 2009)
[a] D=*Dermatophagoides*, Am=*Amblyomma*, As=*Ascoschoengastia,* C=*Carios*, H=*Haemaphysalis*, I=*Ixodes*, Le=*Leptotrombidium*, M= *Metaseiulus*, O=*Ornithodoros*, R=*Rhipicephalus*, *S*=*Steganacarus*, T=*Tetranychus*, U=*Unionicola*, V=*Varroa*, W=*Walchia*, Li=*Limulus*
[b] S= Acari superorder (A = Acariformes, P= Parasitiformes)
[c] PCG= Protein coding genes
[d] duplications of the control region were also considered
[e] Ref=References;(un) = unpublished; (1) Shao *et al*. 2004; (2) Black and Roehrdanz 1998; (3) Shao *et al*. 2005b; (4) Shao *et al*. 2006; (5) Shao *et al*. 2005a; (6) Jeyaprakash and Hoy 2007; (7) Shao *et al*. 2004 (8) Mitani *et al*. 2004;(9) Domes *et al*. 2008; (10) Van Leeuwen *et al*. 2000; (11) Navajas *et al*. 2002; (12) Staton *et al*. 1997; Lavrov *et al*. 2000b
[f] GC- and AT-skew for the strand coding for *cox1*, calculated following (Perna and Kocher 1995)
[g] for these species the largest non-coding region(s) was/were assumed to be the control region(s)
[h] *L. pallidum* has a duplication of *16S-rRNA*; in this table the largest *16S-rRNA* gene is considered
[i] only single copy genes were considered for protein coding gene length calculation of *M. occidentalis*
[j] *L. polyphemus* belongs to the order of the Xiphosura within the class of the Merostomata

Table II.3 - Pairwise common interval distance matrix of mt gene orders of Acari[*].

	Dp	*I*	*Rs*	*Vd*	*La*	*Tu*	*A*	*Wh*	*Uf*	*Sm*
D. pteronyssinus (*Dp*)	**1,326**/*204*	**70**	**74**	**46**	**18**	**14**	**18**	**18**	**16**	**-**
Ixodes sp. (*I*) [a]	*56*	**1,326**/*204*	**388**	**278**	**34**	**26**	**68**	**70**	**78**	**-**
R. sanguineus (*Rs*) [b]	*86*	*78*	**1,326**/*204*	**202**	**36**	**26**	**72**	**70**	**74**	**-**
V. destructor (*Vd*)	*56*	*204*	*78*	**1,326**/*204*	**36**	**18**	**60**	**60**	**54**	**-**
L. akamushi (*La*) [c]	*20*	*36*	*24*	*36*	**1,326**/*204*	**20**	**100**	**98**	**90**	**-**
T. urticae (*Tu*)	*22*	*26*	*22*	*26*	*22*	**1,326**/*204*	**20**	**20**	**16**	**-**
Ascoschoengastia sp.(*A*)	*34*	*38*	*46*	*38*	*44*	*28*	**1,326**/*204*	**278**	**130**	**-**
W. hayashii (*Wh*)	*24*	*48*	*32*	*48*	*32*	*46*	*40*	**1,326**/*204*	**122**	**-**
U. foili (*Uf*)	*30*	*90*	*44*	*90*	*38*	*32*	*40*	*42*	**1,326**/*204*	**-**
S. magnus (*Sm*) [d]	*60*	*132*	*72*	*132*	*28*	*22*	*34*	*62*	*80*	**-** /*204*

*bold numbers represent pairwise common interval distances between mt gene orders of Acari (37 genes in total), while italic numbers represent pairwise common interval distances between mt gene orders of Acari without tRNAs (15 genes in total) (*L. pallidum* and *M. occidentalis* were excluded from the dataset as the genomes contain duplicated genes whereas other genes are absent). Accession numbers of Acari mt genomes are listed in Table II.2.

[a] similar gene order as *L. polyphemus*, *Ornithodoros* sp. and *C. capensis*

[b] similar gene order as *H. flava* and *A. triguttatum*

[c] similar gene order as *L. deliense*

[d] due to lack of tRNAs only protein coding gene and rRNA order comparison was possible

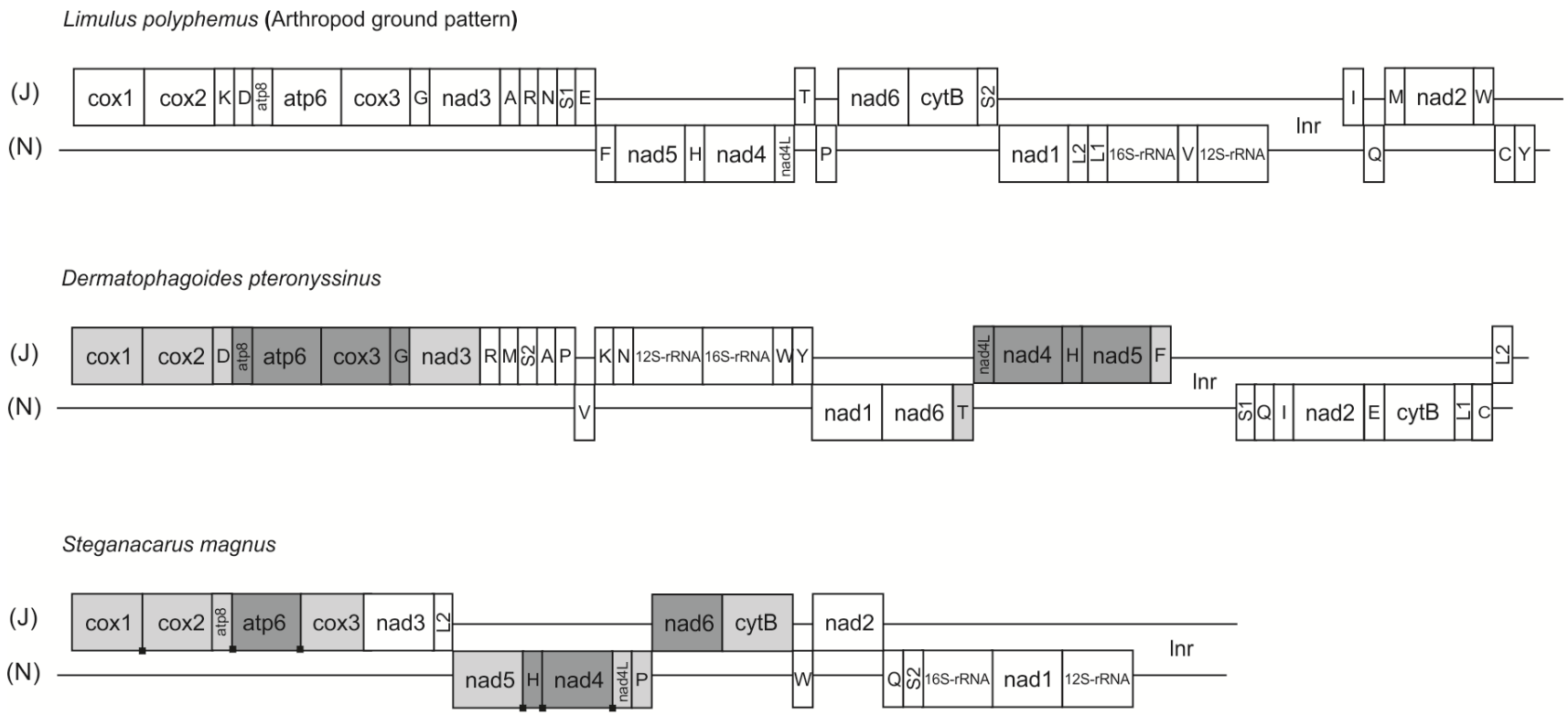


Figure II.4 - Mitochondrial gene arrangement of *Limulus polyphemus*, *Dermatophagoides pteronyssinus* and *Steganacarus magnus*.
Graphical linearisation of mt genomes is presented according to Fahrein *et al.* (2007). Gene sizes are not drawn to scale. J stands for majority and N for minority strand. Protein coding and rRNA genes are abbreviated as in the abbreviations section. tRNA genes are abbreviated using the one-letter amino acid code, with L_1=CUN; L_2=UUR; S_1= AGN; S_2=UCN. Dark grey boxes indicate conserved gene boundaries at the 5'- and 3'-end compared to *L. polyphemus*. Light grey boxes indicate conserved gene boundaries at the 5'- or 3'-end compared to *L. polyphemus*. Square dots between the genes of *S. magnus* represent conserved gene boundaries compared to *D. pteronyssinus*
.

All 37 genes present in a standard metazoan mt genome could be identified (Fig. II.2). Gene overlap exists between *trnD/atp8* (1bp), *trnR*/*nad3* (17bp), *trnM*/*trnS*$_2$ (12bp), *trnP*/*trnV* (11bp), *trnV*/*trnK* (7bp), *trnW*/*trnY* (1bp), *trnY*/*nad1* (4bp), *trnI*/*trnQ* (1bp) and *trnL*$_1$/*trnC* (5bp). No overlap was found between protein coding genes. Small non-coding regions (> 20bp) are present between *trnS*$_2$/*trnA* (25bp), *trnA*/*trnP* (20bp) and *nad1*/*nad6* (28bp). A large non-coding region is positioned between *trnF* and *trnS*$_1$ (286bp). Twenty-five genes of the mt genome of *D. pteronyssinus* are transcribed on the majority strand (J-strand), whereas the others are oriented on the minority strand (N-strand).

The mt genome of the horseshoe crab *Limulus polyphemus* is considered to represent the ground pattern for arthropod mt genomes (Staton *et al.* 1997; Lavrov *et al.* 2000a). Comparing the *D. pteronyssinus* genome to this sequence revealed that only 11 of the 38 gene boundaries in *L. polyphemus* are conserved in *D. pteronyssinus* (Fig. II.4, (Fahrein *et al.* 2007)). Moreover, by making use of the pattern search function in the Mitome-database ((Lee *et al.* 2008), status January 10, 2009), the mt gene order of *D. pteronyssinus* appeared to be unique among arthropods. Remarkably, the relative position of *trnL*$_2$ (between *nad1* and *16S-rRNA*), which differentiates the Chelicerata, Myriapoda and Onychophora from the Insecta and Crustacea according to Boore (1998; 2006a), is not conserved. However, Boore's hypothesis was based on mt genome data from only 2 Chelicerata that were available in 1998. At the time of this study, 41 complete chelicerate mt genomes were available in the NCBI-database (http://www.ncbi.nlm.nih.gov/, status January 10, 2009). Out of these, only 29 depict the specific arrangement of *trnL*$_2$ between *nad1* and *16S-rRNA* (see Appendix II-A for an overview of gene arrangements of chelicerate mt genomes). This illustrates that care should be taken when general rules are deduced from limited datasets.

Mt gene arrangements have already provided strong support toward the resolution of several long-standing controversial phylogenetic relationships (Boore 2006a). Surprisingly, the mt gene order of *D. pteronyssinus* differs considerably from that of other mites (see Appendix II-A). Comparing the *D. pteronyssinus* mt genome to the mt sequence of the oribatid *S. magnus* (Domes *et al.* 2008), the closest relative of *D. pteronyssinus*, revealed that only 6 of the 22 gene boundaries in *S. magnus* are conserved in *D. pteronyssinus* (Fig. II.4). Extending this analysis to the other Acari mt genomes showed that in several cases the set of neighbouring genes that were not separated during the evolution (Berard *et al.* 2007) was greater between members of different superorders (e.g. *D. pteronyssinus* (Acariformes) and *Rhipicephalus sanguineus* (Parasitiformes)) than between members of the same superorder (e.g. *D. pteronyssinus* (Acariformes) and *T. urticae* (Acariformes)) (Table II.3). Exclusion of tRNAs

in our analysis showed a similar trend, suggesting that protein coding genes were also involved in mt gene rearrangements. These results indicate that mt gene orders seem less useful for deduction of phylogenetic relationships between superorders within the Acari. However, comparing gene order might be more powerful to establish phylogenetic relations within families, as was previously proposed (Curole and Kocher 1999; Nardi *et al.* 2001). In the case of the Ixodidae family, it was shown that the division of Prostriata (*Ixodes* sp.) and Metastriata (*R. sanguineus*, *Amblyomma triguttatum*, *Haemaphysalis flava*) could be linked to mt gene arrangements (Black and Roehrdanz 1998; Shao *et al.* 2004).

3.2 Base composition and codon usage

The overall AT-content of the mt genome of *D. pteronyssinus* is 72.6% (Table II.2). This is within the range of the average AT-content of Acari mt genomes (74.6+/- 4.0%). The high AT-content is reflected in the codon usage (Table II.4 (Sharp *et al.* 1988)) with nucleotides 'A' and 'T' preferred over 'C' and 'G' on the wobble position and the predominant use of codons deficient in 'C' or 'G'. For example, the most frequently used codons are TTT (F) (105 codons per 1000 codons) and TTA (L_2) (78 codons per 1000 codons).

Metazoan mt genomes usually present a clear strand bias in nucleotide composition (Hassanin *et al.* 2005). This is probably due to asymmetric patterns of mutations during transcription and replication when one strand remains transiently in a single-stranded state, making it more vulnerable to DNA damage (Francino and Ochman 1997). However, in the case of mtDNA-replication, this hypothesis is not without controversy (Clayton 1991; Yang *et al*. 2002a; Brown *et al*. 2005; Yasukawa *et al*. 2005). The strand bias in nucleotide composition can be measured as GC- and AT-skews ((G%-C%)/(G%+C%) and (A%-T%)/(A%+T%), respectively) (Perna and Kocher 1995). The overall GC- and AT-skews of the J-strand of the *D. pteronyssinus* mt genome are 0.194 and -0.199, respectively. At the time of writing, these were the most extreme values encountered within mite mt genomes (Table II.2) and they are reversed compared to the usual strand biases of metazoan mtDNA (negative GC-skew and positive AT-skew for the J-strand). Moreover, a positive GC-skew for mite mt genomes seems to be rare as it was only encountered in *Varroa destructor* in 2009. Although hypothetical, it could be the result of a strand swap of the control region (Hassanin *et al.* 2005). A strand swap or inversion of the control region is expected to produce a global reversal of asymmetric mutational constraints in the mtDNA, resulting with time in a complete reversal of strand compositional bias (see Chapter I) (Hassanin *et al.* 2005). The asymmetrical directional mutation pressure is also reflected in the codon usage of

genes oriented in opposite directions (Carapelli *et al.* 2008). Whereas NNG and NNU codons are preferred over NNA and NNC codons on the J-strand, genes on the N-strand show the exact opposite trend (see Fig. II.5 for an across-strand (N and J) comparison of frequencies of codons ending with the same nucleotide).

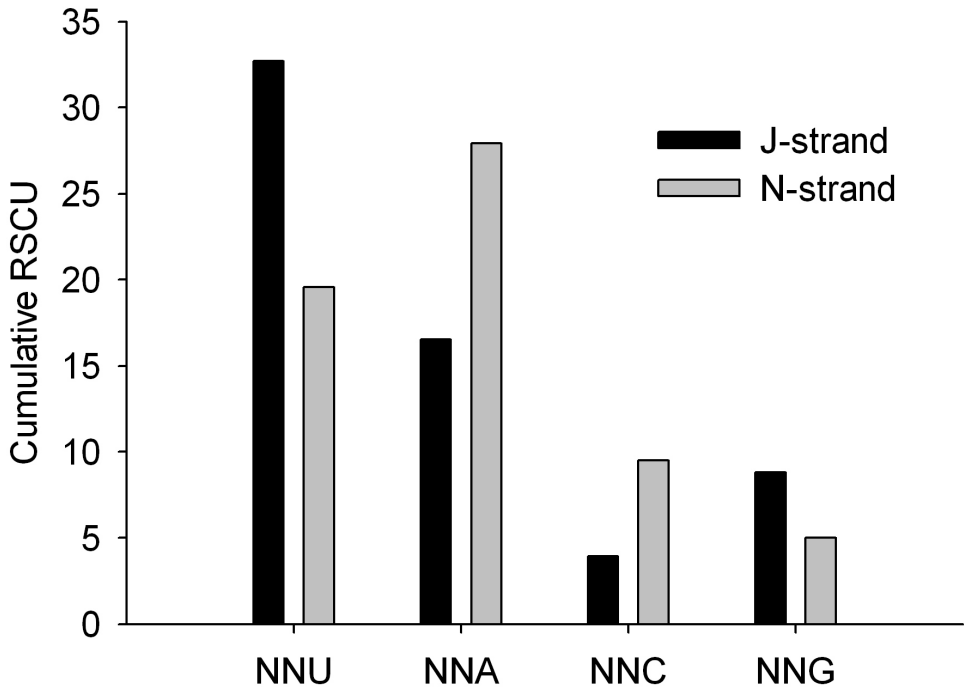


Figure II.5 - Across-strand (N and J) comparison of frequencies of codons ending with the same nucleotide

Values on the y-axis represent the sum of Relative Synonymous Codon Usage (RSCU) values (Table II.4) of codons ending with the same nucleotide across all codon families (x-axis).

Table II.4 - Relative synonymous codon usage (RSCU) and number of codons per 1000 codons (NC1000) in the protein coding genes of the mitochondrial genome of *D. pteronyssinus*.

Amino acid	codon	RSCU*	NC1000	Amino acid	codon	RSCU	NC1000	Amino acid	codon	RSCU	NC1000	Amino acid	codon	RSCU	NC1000
F	TTC	0.24	14.14	**S2**	TCA	1.13	15.52	**Y**	TAC	0.48	11.09	**C**	TGC	0.33	2.77
	TTT	1.76	105.60		TCC	0.43	5.82		TAT	1.52	35.20		TGT	1.67	14.14
L2	TTA	3.43	78.71		TCG	0.14	1.94					**W**	TGA	1.11	12.75
	TTG	0.81	18.57		TCT	3.81	52.11						TGG	0.89	10.25
L1	CTA	0.74	16.91	**P**	CCA	1.53	14.41	**H**	CAC	0.50	4.16	**R**	CGA	1.37	3.88
	CTC	0.10	2.22		CCC	0.35	3.33		CAT	1.50	12.47		CGC	0.00	0.00
	CTG	0.10	2.22		CCG	0.24	2.22	**Q**	CAA	1.68	8.59		CGG	0.49	1.39
	CTT	0.83	19.12		CCT	1.88	17.74		CAG	0.32	1.66		CGT	2.15	6.10
I	ATC	0.40	16.08	**T**	ACA	1.33	13.58	**N**	AAC	0.63	10.53	**S1**	AGA	0.99	13.58
	ATT	1.60	64.30		ACC	0.44	4.43		AAT	1.37	23.00		AGC	0.16	2.22
M	ATA	1.58	57.65		ACG	0.05	0.55	**K**	AAA	1.67	28.55		AGG	0.61	8.31
	ATG	0.42	15.52		ACT	2.18	22.17		AAG	0.33	5.54		AGT	0.73	9.98
V	GTA	1.09	21.06	**A**	GCA	0.71	4.99	**D**	GAC	0.69	6.93	**G**	GGA	0.96	13.86
	GTC	0.17	3.33		GCC	0.59	4.16		GAT	1.31	13.03		GGC	0.15	2.22
	GTG	0.43	8.31		GCG	0.16	1.11	**E**	GAA	0.95	10.81		GGG	1.45	21.06
	GTT	2.31	44.62		GCT	2.53	17.74		GAG	1.05	11.92		GGT	1.44	20.79

* RSCU is the number of times a particular codon is observed relative to the number of times a codon would be observed in the absence of any codon usage bias (Sharp *et al.* 1988).

3.3 Protein coding genes

Nine proteins are encoded by genes on the J-strand (*cox1*, *cox2*, *cox3*, *atp6*, *atp8*, *nad3*, *nad4*, *nad4L*, *nad5*), while four are encoded by genes on the N-strand (*nad1, nad6, nad2, cytB*). The total length (10,826bp) and AT-content (71.61%) of the protein-coding genes are within the range of values typical for Acari (10,639.0 +/- 272.0bp; 74.0 +/- 4.0%, respectively) (Table II.2). Compared to other mite mt proteins, cox1, cox2 and cytB are best conserved. On the other hand, atp8, nad6 and nad4L showed lowest similarity values (see Fig. II.6 for the average identity and similarity % of mt proteins of *D. pteronyssinus*).

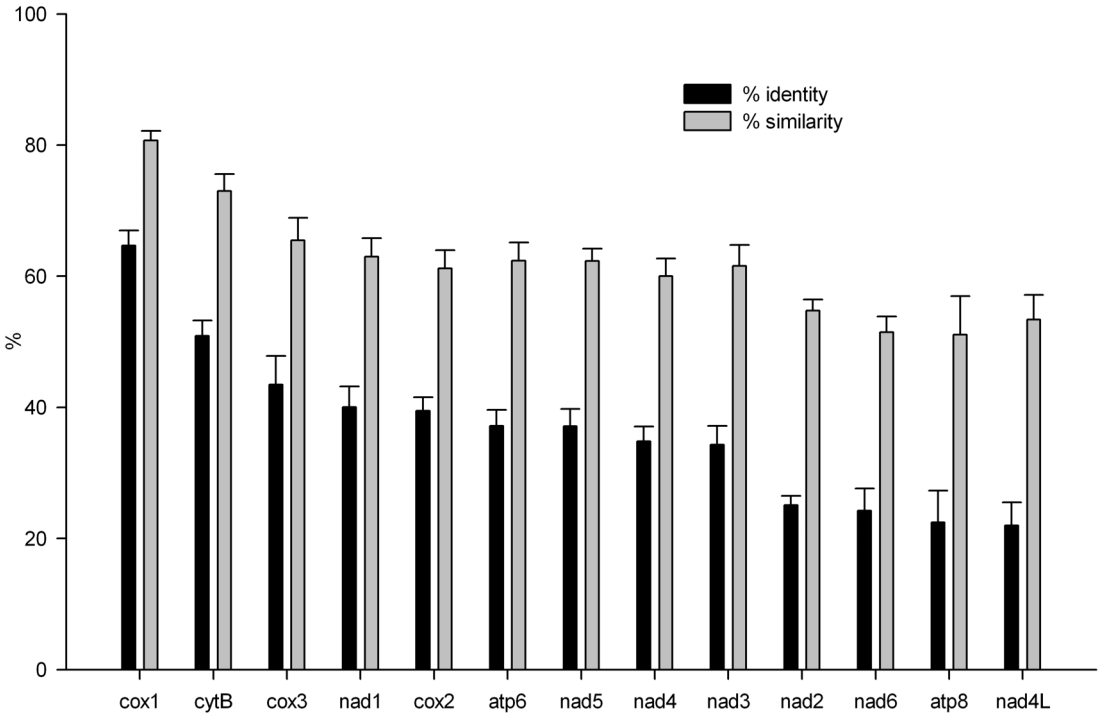


Figure II.6 - Average identity and similarity % of mt proteins of *D. pteronyssinus*
For each protein of *D. pteronyssinus*, a similarity and identity value was calculated with the corresponding protein of other Acari species (as listed in Table II.2), using pairwise global alignment. The obtained values were used to calculate an average identity and similarity % for each protein.

Start and stop codons were determined based on alignments with the corresponding genes and proteins of other mite species. In the case of stop codons, we could also benefit from available expressed sequence tags (ESTs) of *D. pteronyssinus* (n=1,797) and *D. farinae* (n=1,735) (Fig. II.2) (Gissi and Pesole 2003). As for other metazoan mt proteins, unorthodox initiation codons are used (Wolstenholme 1992) (see Appendix II-B for start and stop codons

of protein coding genes of Acari mt genomes). Eight genes (*cox2, atp6, cox3, nad3, nad6, nad4L, nad4, cytB*) use the standard ATG start codon, 3 genes (*cox1, nad1, nad2*) start with ATA and *nad5* initiates with ATT. *atp8* most likely starts with codon TTG.

Eleven genes employ a complete translation termination codon, either TAG (*cox1, cox3*) or TAA (*cox2, atp8, atp6, nad1, nad3, nad6, nad4L, nad5, cytB*). With the exception of *nad3*, *atp8* and *nad4L*, *D. pteronyssinus* ESTs for all these genes confirmed the position of the stop codon (Fig. II.2, see Appendix II-C for sequences of *Dermatophagoides* ESTs covering the mt genome of *D. pteronyssinus*). Berthier *et al.* (1986) showed that the adjacent genes, *nad4L*/*nad4* and *atp8*/*atp6,* were transcribed and translated as a bicistronic mRNA in the model organism *Drosophila melanogaster*. However, as no ESTs were found that aligned with the *nad4L/nad4* and *atp8/atp6* gene boundaries, it could not be confirmed whether this was also the case for *D. pteronyssinus*. Despite its efficiency, the use of sequence alignments to determine the position of stop codons resulted in several cases in overlapping genes. For example, based on a highly conserved tryptophan at the C-terminal end of Acari nad3 proteins, a stop codon was positioned despite the resulting 17bp overlap with *trnaR*. The two remaining genes (*nad2* and *nad4*) are likely equipped with a truncated stop codon (T). Polyadenylation of the mRNA is needed in these cases to form a fully functional TAA stop codon (Ojala *et al.* 1981). Although speculative, ESTs of *D. farinae* confirm the truncated stop of *nad4* (Fig. II.2, see Appendix II-C).

3.4 Transfer RNAs

Fourteen tRNAs are encoded on the J-strand and 8 on the N-strand (Fig. II.2). Secondary structures were predicted for all tRNAs (Fig. II.7). With the exception of $trnS_1$ (UCU instead of GCU in *L. pallidum*) and *trnP* (UGG instead of AGG in *S. magnus*), all anticodon sequences were identical to those of *L. pallidum* and *S. magnus*, the only acariform mites for which tRNA secondary structures have been reported (Shao *et al*. 2005a; Domes *et al*. 2008). Usually, T is in the first anticodon position for tRNAs that recognise either four-fold degenerate codon families or NNR codons. G is usually in this position only to specifically recognize NNY-codons (Boore 2006b). Except for *trnM*, all of the *D. pteronyssinus* mt tRNAs follow this pattern. *trnM* has the anticodon CAT (to recognise both ATG and ATA), which is the case for almost all animal mt systems (Boore 2006b) (Fig. II.7).

Only one tRNA lacks the D-arm: $trnS_1$, as is common for most metazoans. With the exception of *trnC*, *trnV* and $trnS_1$, all tRNAs have T-arm variable loops (TV replacement loops) instead of the T-arm. Similar structures were found for tRNAs of *L. pallidum* (Shao *et*

al. 2005a) and *S. magnus* (Domes *et al.* 2008). The absence of the T-arm is a typical feature for tRNAs of Chelicerata belonging to the orders of the Araneae, Scorpiones and Thelyphonida. However, other taxa within the Chelicerata (Amblypygi, Opiliones, Ricinulei, Solifugae and ticks) possess typical metazoan cloverleaf tRNAs (Masta and Boore 2008). Masta and Boore (2008) suggested a multi-step evolutionary process in an attempt to understand how so many tRNAs in these chelicerate groups could lose their T-arm. According to this speculative theory, changes in mt ribosomes, resulting in the fact that the loss of arms from tRNAs was tolerated (Wolstenholme *et al.* 1987; Masta 2000), and/or changes in specific elongation factors (Ohtsuki *et al.* 2001; Arita *et al.* 2006; Ohtsuki and Watanabe 2007) are considered as a first step in this process.

Only 7 of the 22 tRNAs have a completely matched 7bp acceptor stem (*trnG, trnW, trnH, trnF, trnE*, $trnL_1$ and $trnL_2$). A maximum of 3 mismatches in this stem is found in *trnR*. In contrast, almost all tRNAs (18) possess a completely matched 5bp anticodon stem. *trnC,* $trnS_1$ and *trnN* have a single mismatch whereas *trnY* has two mismatches in this stem. All tRNAs, except $trnL_2$, have a symmetric anticodon loop consisting of 2bp up- and 2bp downstream of the 3bp anticodon. The anticodon loop of $trnL_2$ consists of 2 nucleotides preceding the anticodon and 3 nucleotides immediately following it. This kind of aberrant anticodon loops have also been reported for the two-humped camel *Camelus bactrianus ferus* ($trnS_2$) (Cui *et al.* 2007) and the scorpion *Mesobuthus gibbosus* (*trnH* and *trnN*) (Davila *et al.* 2005). As mentioned before, sequences of some tRNAs overlap with neighbouring genes. The extreme examples are *trnR,* $trnS_2$ and *trnV*. *trnR* overlaps with the adjacent gene *nad3* on the same strand for 17bp at its 3'-end whereas $trnS_2$ overlaps with the adjacent gene *trnM* on the same strand for 12bp at its 3'-end. *trnV* overlaps with the adjacent gene *trnP* on the opposite strand for 11bp at its 3'-end and with *trnK* on the opposite strand for 7bp at its 5'-start. Despite these overlaps, we consider these genes not likely to be pseudogenes. First of all, their sequence is relatively well conserved when compared to corresponding genes of other Acari. Secondly, besides sequence conservation they depict a conserved secondary structure. Thirdly, an EST (GenBank accession number: CB284825) of the related species *D. farinae* was found corresponding to the region covering *trnR*, *trnM* and $trnS_2$ of *D. pteronyssinus* indicating that the genes are expressed (see Fig. II.8 for an alignment of *trnR*, *trnM* and $trnS_2$ of *D. pteronyssinus* with an EST of *D. farinae*). Finally, and most importantly, stem mismatches and sequence overlap are not uncommon for mt tRNAs of arachnids (2004; 2008), and are probably repaired by a post-transcriptional editing process (Lavrov *et al.* 2000b; Masta 2000).

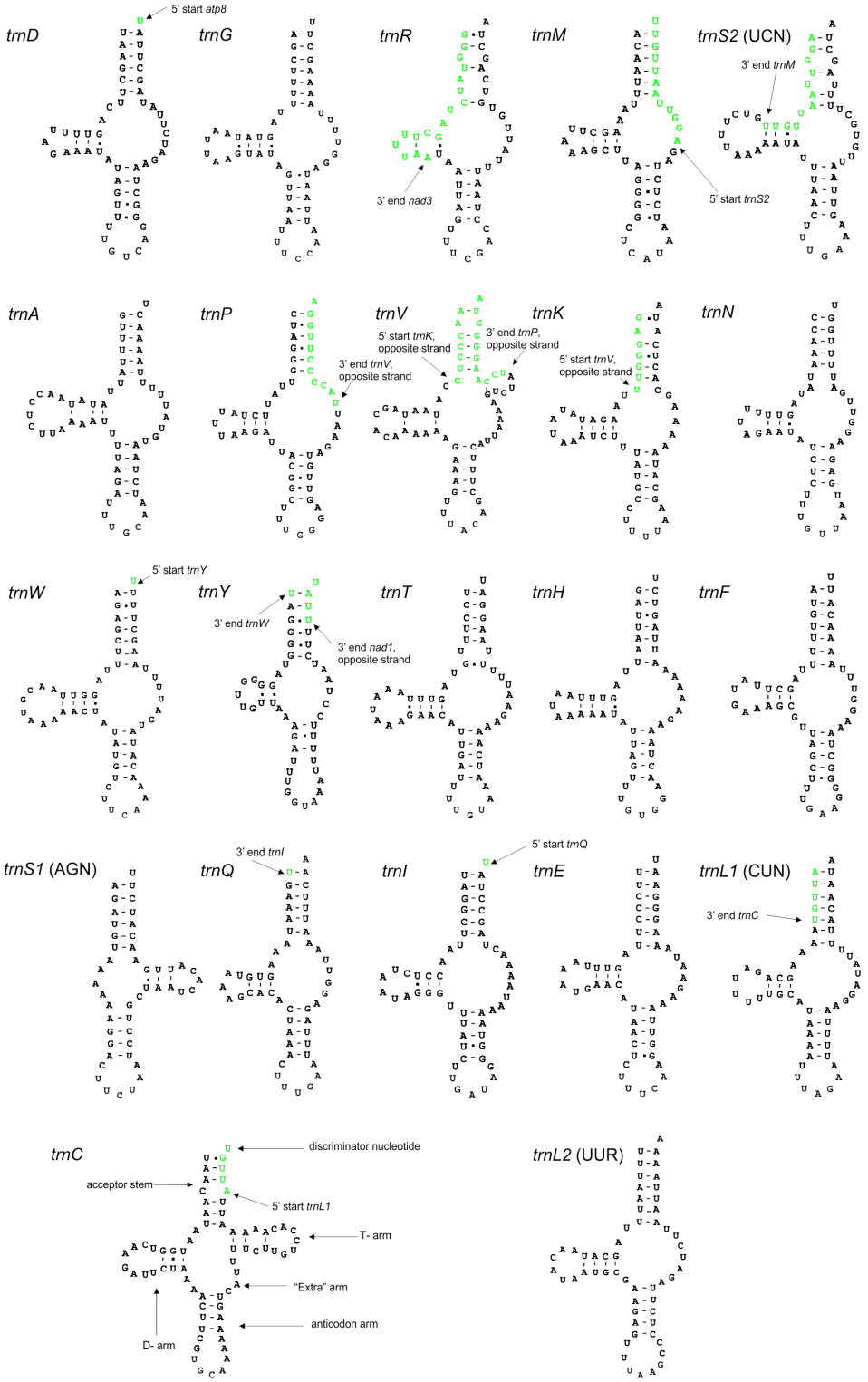


Figure II.7 - Inferred secondary structures of the 22 mitochondrial tRNAs from *D. pteronyssinus*.

tRNAs are shown in the order of occurrence in the mt genome starting from *cox1*. Locations of adjacent gene boundaries are indicated with arrows. Green font indicates that the sequence is part of the adjacent gene. Inferred Watson-Crick bonds are illustrated by lines, whereas GU bonds are illustrated by dots.

Figure II.8 - Alignment of a mt genome fragment containing *trnaM*, *trnaR* and *trnaS*$_2$ of *D. pteronyssinus* and an EST [GenBank: CB284825] of *D. farinae*

Description: Anticodons and anticodon stems are red and green respectively. Acceptor-stems of *trnR, trnM* and *trnS*$_2$ are underlined. The stop codon of *nad3* is grey shaded.

3.5 Non-coding regions

The largest non-coding region (286bp) is flanked by *trnF* and *trnS$_1$*. It is highly enriched in AT (91.61%) and can form stable stem-loop secondary structures. Based on these features, it possibly functions as a control region (Wolstenholme 1992; Zhang and Hewitt 1997). With the exception of *T. urticae* (95.45%), it has the highest AT-content of all Acari mt control regions (Table II.2). The position of the non-coding region differs from most insect and arachnid mt genomes, where the region is mostly located in close proximity to *12S-rRNA* ((Zhang and Hewitt 1997), see Appendix II-A).

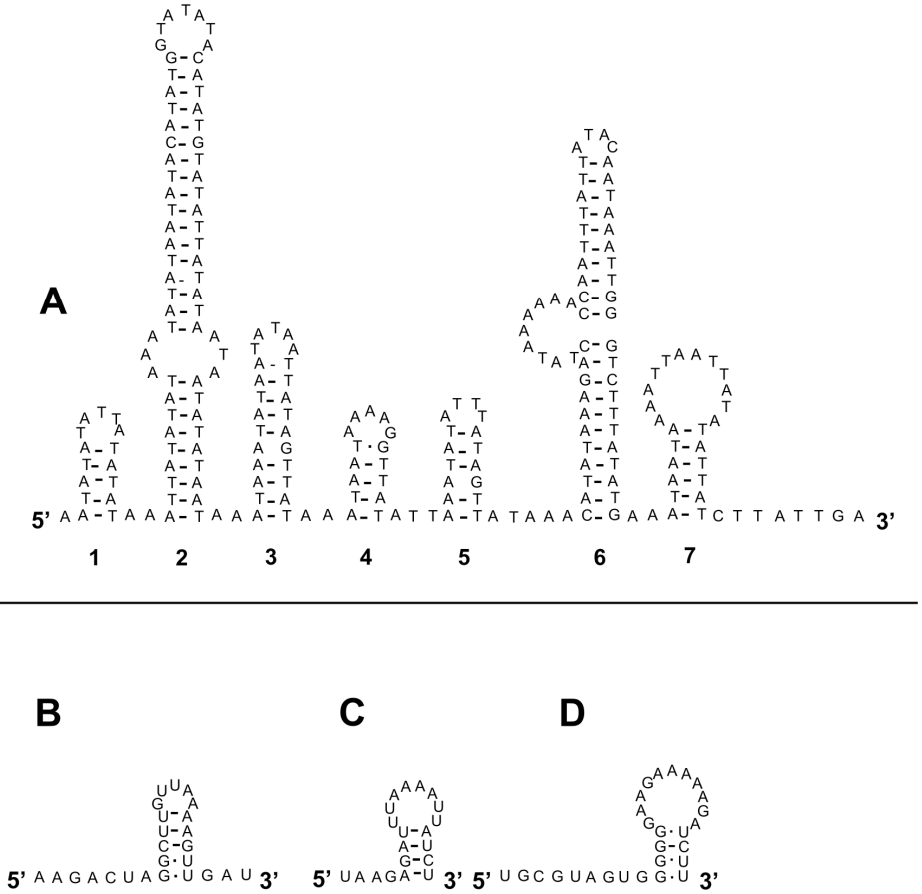


Figure II.9 - Secondary structures of non-coding regions of the mt genome of *D. pteronyssinus*.

Secondary structure of non-coding regions between (A) *trnF* and *trnS$_1$* (large non-coding region); (B) *trnS$_2$* and *trnA;* (C) *trnA* and *trnP;* (D) *nad1* and *nad6*. All structures were constructed using Mfold (Zuker 2003). Inferred Watson-Crick bonds are illustrated by lines, whereas GU bonds are illustrated by dots.

Based on the sequence pattern, the control region can be subdivided in a repeat region and a stem-loop region. The first region (11,491-11,528bp) contains several AT-repeats. In order to verify the exact number of repeats we resequenced this region. For this purpose, two flanking primers, Dp-Ms-F and Dp-Ms-R, were synthesised spanning approximately 700bp. The PCR product was cloned and ten independent clones were sequenced. This revealed that the number of AT-repeats varied between 7 to 28, suggesting that this domain can be considered as a microsatellite (Goldstein and Schlotterer 1999). This is remarkable as a mt microsatellite was never reported before for species belonging to the Chelicerata. Also in metazoan mtDNA such microsatellites are rare and have, to our knowledge, only been reported for butterflies (Cameron and Whiting 2008), a dragonfish, *Scleropages formosus* (Yue *et al.* 2006), a bat, *Myotis bechsteinii* (Mayer and Kerth 2005), a turtle, *Pelomedusa subrufa* (Zardoya and Meyer 1998) and several seal species (Arnason and Johnsson 1992; Arnason *et al.* 1993; Hoelzel *et al.* 1993).

The second region (11,529-11,768bp) holds two short palindromic sequences, TACAT and ATGTA, which are conserved in mt genomes of mammals (Saccone *et al.* 1991) and fishes (Zardoya and Meyer 1996; Yue *et al.* 2006). They can form a stable stem-loop structure (Fig. II.9*A2*), which might be involved as a recognition site for the arrest of J-strand synthesis (Saccone *et al.* 1991). Near this region other stem-loop structures could be folded (Fig. II.9*A*) but none of them had flanking sequences similar to those that are conserved in the control region of the mt genome of insects (Zhang and Hewitt 1997) and metastriate ticks (Black and Roehrdanz 1998).

As described before, four other stretches of non-coding nucleotides were found outside the control region. These short sequences can fold into stable stem-loop structures (Fig. II.9*B,C*, and *D*) which may function as splicing recognition sites during processing of the transcripts (He *et al.* 2005).

3.6 Ribosomal RNAs

12S-rRNA and *16S-rRNA* are located on the J-strand. This does not coincide with their position in most Chelicerata where they are located on the N-strand (see Appendix II-A). The AT-contents of both genes are comparable (72.9% and 76.1% for the *12S-* and *16S-RNA*, respectively) and are within the range of rRNAs of other Acari (76.5 +/- 4.2%; 78.0 +/- 4.1%, respectively). The sizes of the rRNAs (665bp and 1,078bp) are slightly larger than those of other acariform mite rRNAs (626.0 +/-29.9bp and 1,018.5 +/-21.3bp) but are shorter than those found in the Parasitiformes (706.0 +/- 17.5bp and 1,207.3 +/-31.4bp) (Table II.2).

The *12S-rRNA* and *16S-rRNA* genes of *Leptotrombidium* species (Acariformes: Trombidiformes: Trombiculidae) are 23.4% and 23.5% shorter than their counterparts in *Drosophila yakuba*. This substantial reduction is mainly caused by the loss of stem-loop structures at the 5'-end of the rRNA genes (Shao *et al.* 2006). To identify whether similar domains are absent in the rRNAs of *D. pteronyssinus*, we constructed their secondary structures (Fig. II.10 (Van de Peer *et al.* 1998; De Rijk *et al.* 1999)). This revealed that the *D. pteronyssinus 12S-rRNA* indeed lacks similar stem-loops as *L. pallidum*, compared to *D. yakuba*. The structure also revealed 1 additional stem-loop (stem-loop 1) not present in *12S-rRNA* of *L. pallidum*. Like in *L. pallidum*, one stem-loop replaces three stem-loops (24, 25 and 26) whereas another replaces a region of four stem-loops (39, 40, 41 and 42) of the *D. yakuba 12S-rRNA* (Shao *et al.* 2006). Based on the modelled structure in combination with an alignment of other acariform *12S-rRNAs*, the greatest sequence conservation was found in the loop region of stem-loops 21 and 27 and the region between stem-loops 48 and 50.

In analogy to the *16S-rRNA* gene of *L. pallidum,* the main deletions of the *D. pteronyssinus 16S-rRNA* are located at the 5'-end. With the exception of D19, all stem-loops of *L. pallidum* are present in *D. pteronyssinus*. We also discovered three additional stem-loops (C1, E2 and E19) which are absent in the *16S-rRNA* of *L. pallidum*. The 3'-end of the *16S-rRNA* structure is best conserved compared to other acariform *16S-rRNAs*. This is in agreement with the idea that this region is the main component of the peptidyl-transferase centre, and as such most vulnerable to mutations (He *et al.* 2005). Recently, the *12S-rRNA* and *16S-rRNA* secondary structures of *S. magnus* have been published (Domes *et al.* 2008). The *12S-rRNA* structure of *S. magnus* has 5 extra stem-loops (2, 4, 5, 40 and 42) compared to the one of *D. pteronyssinus* whereas the *16S-rRNA* lacks 6 stem-loops (D4, D16, E1, E2, E19 and G9) and has 5 stem-loops (cd1, D1, D17, D19, G13) not present in the *16S-rRNA* of *D. pteronyssinus.*

It is still an open question how relatively well-conserved structures such as rRNAs can dramatically decrease in size while remaining functional. Wolstenholme *et al.* (1987) and Masta (2000) suggested a correlation between the occurrence of truncated rRNAs (compared to *Drosophila*) and the loss of the T-arm in tRNAs. The coincidence of short rRNAs and missing T-arms in tRNAs was also observed in *S. magnus, L. pallidum* and *D. pteronyssinus*. Other acariform mites like *T. urticae*, *Ascoschoengastia* sp. and *Walchia hayashii* also exhibit short rRNAs (Table II.2) and the prediction of their tRNA secondary structures could further support this hypothesis. However, examples contradicting this hypothesis also occur e.g. pulmonate gastropods with tRNAs lacking T-arms have no truncated rRNAs. Therefore, it

remains possible that truncation of both tRNAs and rRNA genes only reflects an independent trend towards minimisation of the mt genome as suggested by Yamazaki *et al*. (1997).

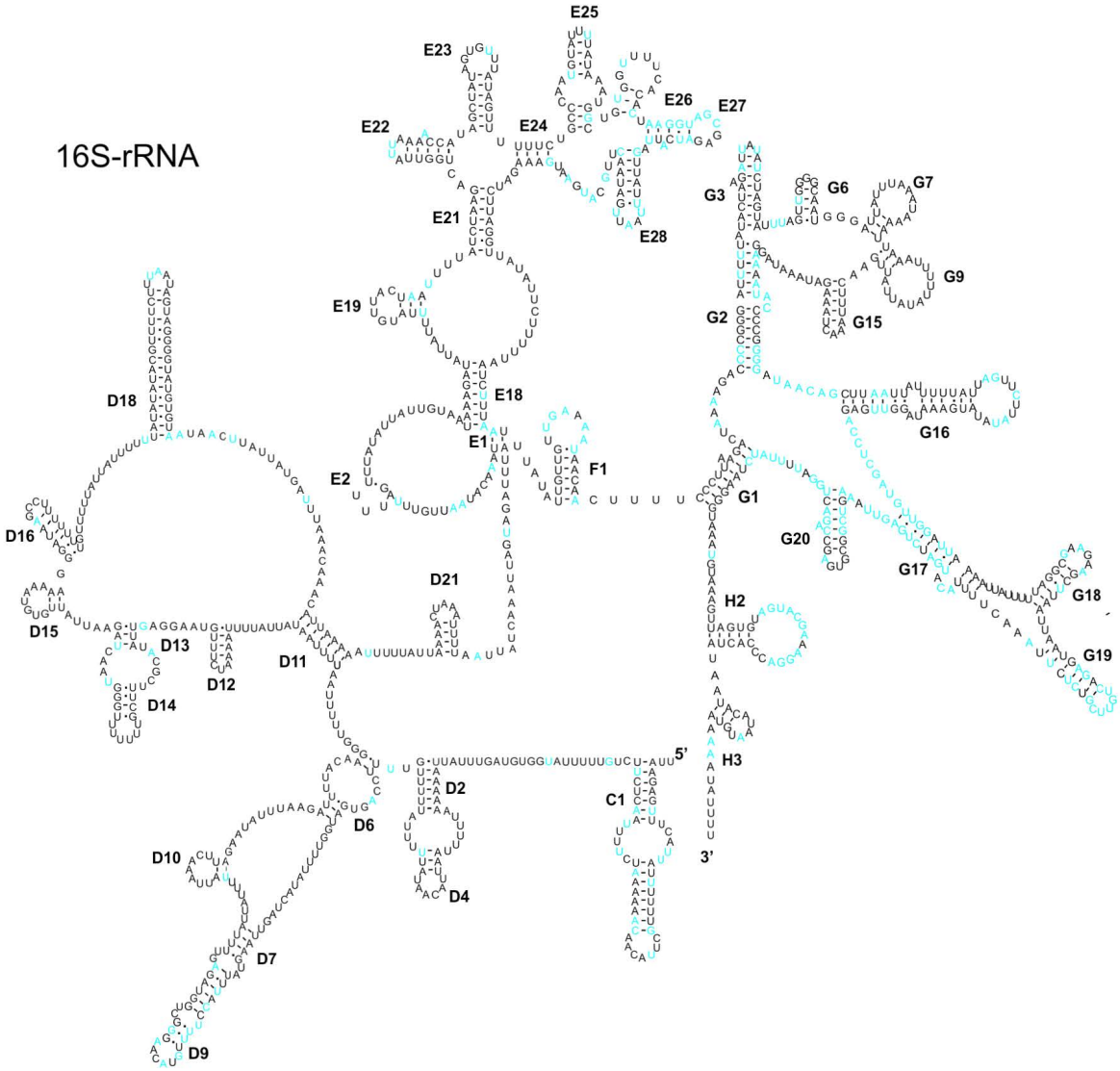


Figure II.10

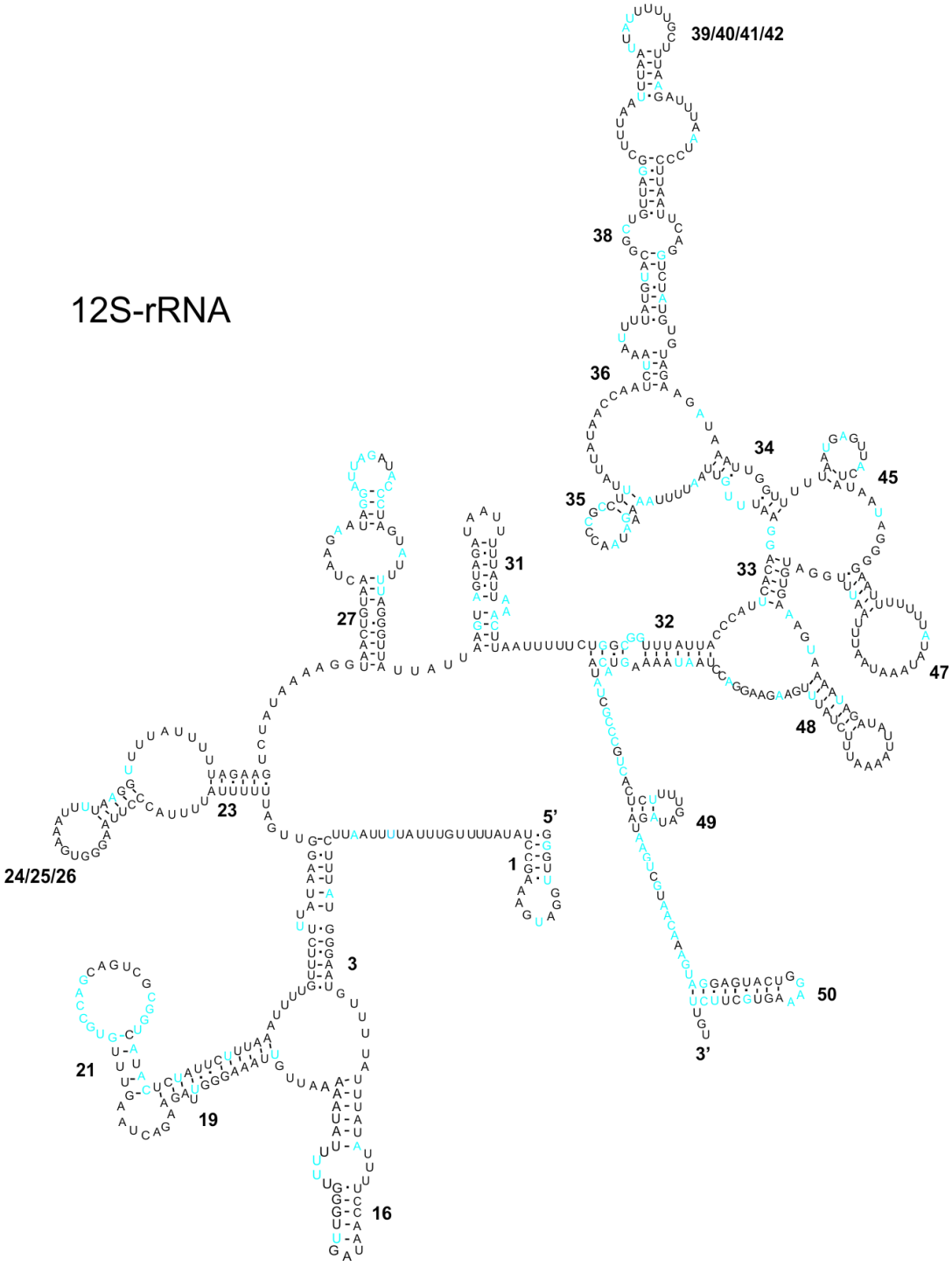


Figure II.10 (continued) – *16S-rRNA* and *12S-rRNA* secondary structure of the mitochondrial genome of *D. pteronyssinus*.

The numbering of the stem-loops is after de Rijk *et al*. (1999) for *16S-rRNA* and after van de Peer *et al*. (1998) for *12S-rRNA*. Blue coloured nucleotides show 100% identity when aligned to *12S-rRNA* and *16S-rRNA* genes from other Acariformes. Inferred Watson-Crick bonds are illustrated by lines, whereas GU bonds are illustrated by dots.

3.7 Phylogenetic analysis

A phylogenetic tree was constructed based on nucleotide and amino acid sequences from all mt protein coding genes of Acari. The ILD-test (Farris *et al.* 1995) indicated a significant incongruence (P=0.01) among data set partitions for nucleotide alignments and low congruence (P=0.07) among data set partitions for amino acid alignments. A considerable debate exists on the utility of this test (Cunningham 1997; Dolphin *et al.* 2000; Yoder *et al.* 2001; Barker and Lutzoni 2002; Dowton and Austin 2002; Hipp *et al.* 2004). However, the principle of Kluge (1989) implies that all data should always be included in a combined analysis for any phylogenetic problem and therefore we combined data partitions for both amino acid and nucleotide alignments for phylogenetic analysis. A maximum parsimony (MP) analysis based on nucleotide alignments (data not shown) grouped *V. destructor* (Parasitiformes) within the Acariformes, close to *D. pteronyssinus*. This is in contrast with the generally accepted view on the phylogeny of the Acariformes and Parasitiformes (Walter *et al.* 1996; Dunlop and Alberti 2008). As mentioned before, *V. destructor* and *D. pteronyssinus* both have a reversal of asymmetrical mutation pattern. When such reversals occurred independently, *D. pteronyssinus* and *V. destructor* could have acquired a similar base composition and as a consequence group together due to the long-branch attraction (LBA) phenomenon (Felsenstein 1978; Hassanin *et al.* 2005). Model-based methods such as maximum likelihood (ML) and Bayesian inference (BI) are less sensitive to LBA (Cunningham *et al.* 1998; Hassanin *et al.* 2005) and were for this reason considered for phylogenetic analysis. ML and BI analysis performed on the amino acid data set resolved trees with an identical topology (Fig. II.11*A*) in which *D. pteronyssinus* clusters with *S. magnus*, forming a sistergroup of the Trombidiformes. This is in agreement with the most recent views on the classification of the Acariformes (O'Connor 1984; Walter *et al.* 1996; Dunlop and Alberti 2008). The nucleotide data set resulted in similar trees, confirming the evolutionary position of *D. pteronyssinus* (Fig. II.11*B*). The only major inconsistency over the trees was the position of *T. urticae*. Although this species is generally considered as a member of the Trombidiformes (O'Connor 1984; 1996; Dunlop and Alberti 2008), it was clustered with the sarcoptiform mites *D. pteronyssinus* and *S. magnus* in the trees based on the nucleotide dataset. (Fig. II.11*B*). However, the position in the different trees is questionable as it is supported by low bootstrap values/Bayesian posterior probabilities (Fig. II.11*A/B*). Adding additional mt genome data from closely related taxa of *T. urticae* and from

taxa located between *T. urticae* and Trombiculidae would probably position *T. urticae* with higher support values within the Trombidiformes.

In the trees based on the nucleotide dataset, *H. flava* is, compared to *A. triguttatum*, evolutionary closer related to *R. sanguineus* while in the trees based on the amino acid dataset this is the opposite. However, as the clustering of *H. flava* and *R. sanguineus* is in agreement with the most recent views on the classification of the Ixodida (Barker and Murrell 2002; Klompen *et al.* 2007), we consider the nucleotide topology as the most correct one. Murrell *et al.* (2005) considers the Parasitiformes to be paraphyletic with respect to the Opilioacariformes, but as there are no complete mt genomes of Opilioacariformes available, we were not able to verify this hypothesis.

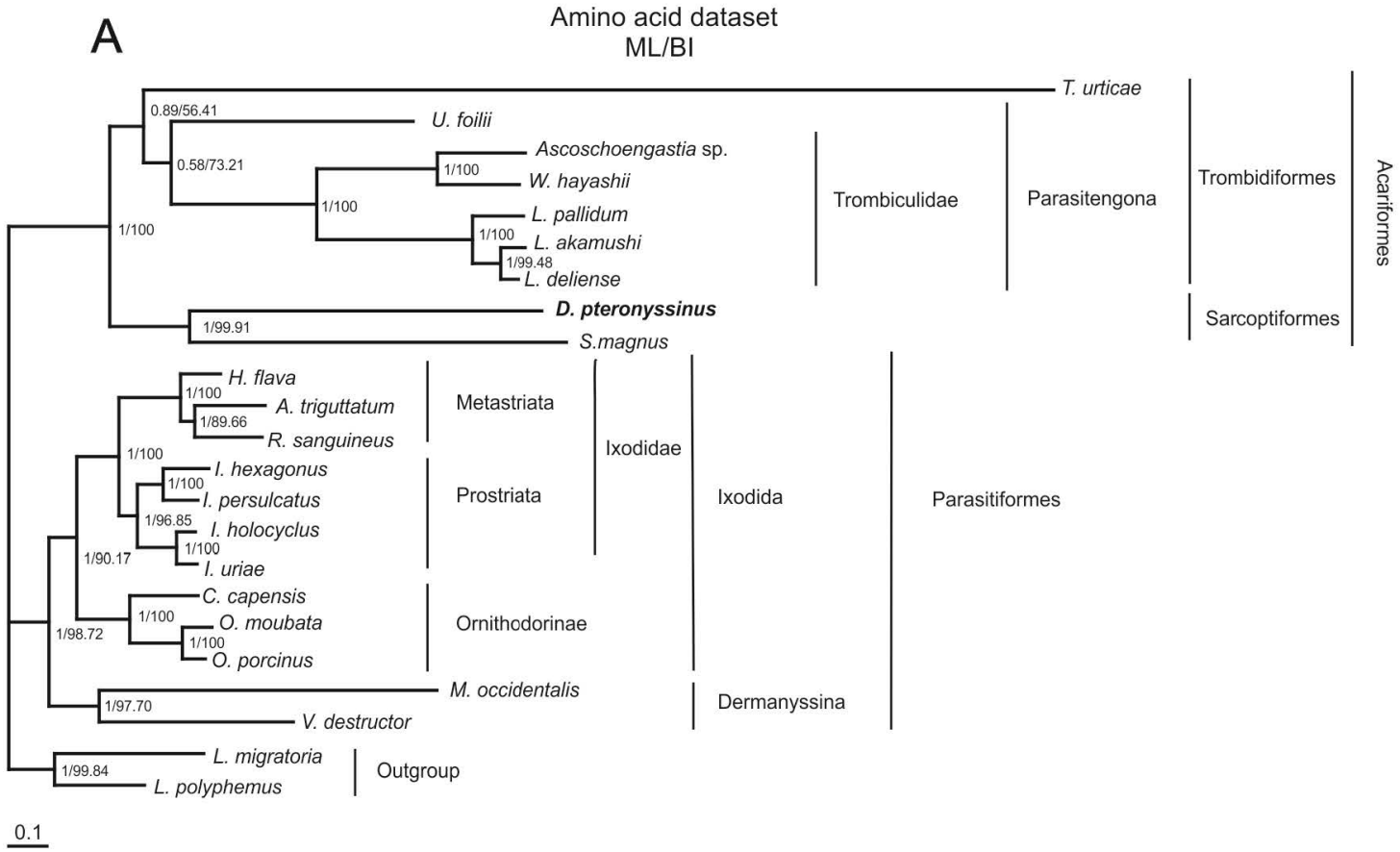

A
Amino acid dataset
ML/BI
T. urticae
U. foilii
Ascoschoengastia sp.
W. hayashii
L. pallidum
L. akamushi
L. deliense
D. pteronyssinus
S.magnus
H. flava
A. triguttatum
R. sanguineus
I. hexagonus
I. persulcatus
I. holocyclus
I. uriae
C. capensis
O. moubata
O. porcinus
M. occidentalis
V. destructor
L. migratoria
L. polyphemus
0.89/56.41
0.58/73.21
1/100
1/100
1/100
1/100
1/99.48
1/99.91
1/100
1/89.66
1/100
1/100
1/96.85
1/100
1/90.17
1/100
1/98.72
1/100
1/97.70
1/99.84
Trombiculidae
Parasitengona
Trombidiformes
Sarcoptiformes
Acariformes
Metastriata
Prostriata
Ixodidae
Ixodida
Parasitiformes
Ornithodorinae
Dermanyssina
Outgroup
0.1

Figure II.11

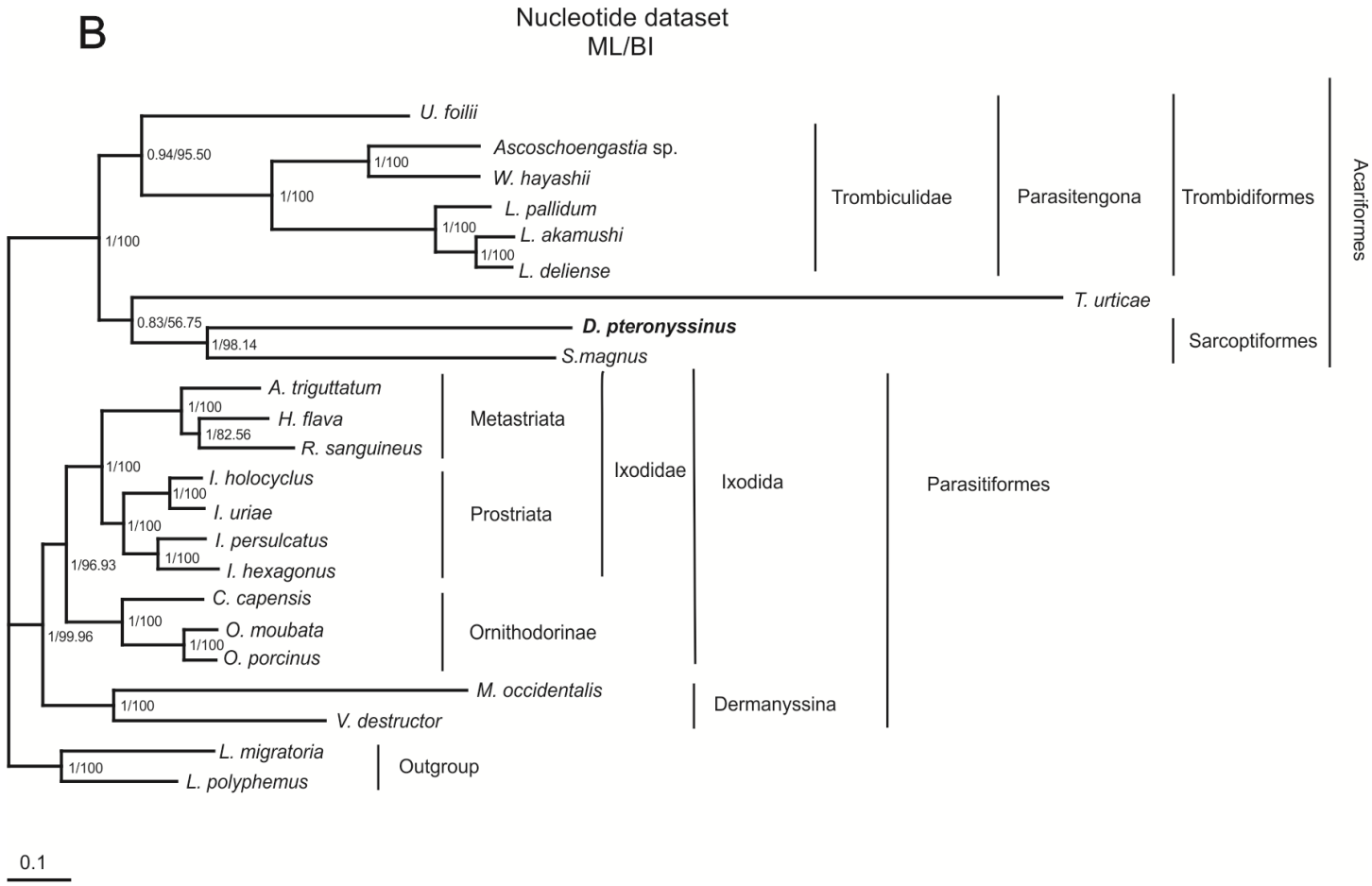


Figure II.11 (continued) - Phylogenetic trees of Acari relationships.

Trees were inferred from amino acid (A) and nucleotide (B) datasets. All protein coding gene sequences were aligned and concatenated; ambiguously aligned regions were omitted by Gblocks 0.91b (Castresana 2000). Trees were rooted with two outgroup taxa (*L. polyphemus* and *L. migratoria*). Numbers behind the branching points are percentages from Bayesian posterior probabilities (left) and ML bootstrapping (right). Accession numbers for the different Acari mt genomes are listed in Table II.2.

4 Conclusions

This study describes the complete mt genome of a species belonging to the Astigmata, a cohort within the Sarcoptiformes. Although the length, gene and AT-content are similar to other Acari mtDNA, the mt genome of *D. pteronyssinus* exhibits some interesting features. The gene order of *D. pteronyssinus* is completely different from that of other Acari mt genomes. Gene order comparison indicated that mt gene orders seem less useful for deduction of phylogenetic relationships between superorders within the Acari. GC- and AT-skews of the J-strand were very large and reversed as compared to those found in most metazoan mtDNA.

Compared to parasitiform mites, both *D. pteronyssinus* rRNAs were considerably shorter and almost all transfer RNAs lacked the T-arm. It would be interesting to investigate whether the occurrence of truncated rRNAs and the loss of the T-arm in tRNAs are correlated or just a trend toward minimisation of the mt genome. Finally, phylogenetic analysis using concatenated mt gene sequences succeeded in recovering Acari relationships concordant with traditional views of phylogeny of Acari.

Chapter III

Mitochondrial genome analysis of the predatory mite *Phytoseiulus persimilis* and a revisit of the *Metaseiulus occidentalis* mitochondrial genome

---

This chapter has been redrafted from:

Dermauw W., Vanholme B., Van Leeuwen T., Tirry L. (2012). Mitochondrial genome analysis of the predatory mite *Phytoseiulus persimilis* and a revisit of the *Metaseiulus occidentalis* mitochondrial genome. Genome 53, 285-301.

1 Introduction

Animal mitochondrial (mt) genomes are typical 14-17kb in size and encode for 13 proteins, 2 ribosomal RNAs and 22 transfer RNAs. They also contain one large non-coding region that fulfils a role in the initiation of replication and transcription (Boore 1999; Taanman 1999). Their relatively small size, extreme compactness and non-Mendelian inheritance made them an important research topic for different scientific disciplines including comparative genomics, phylogenetics and population genetics. Recently it was shown how mutations in the mt genome of the two-spotted spider mite (*Tetranychus urticae*) provided resistance against the pesticide bifenazate (Van Leeuwen *et al*. 2008). The resulting rapid spread of this type of maternally inherited resistance in the field, illustrated the importance of mt genomes in pest management.

Owing to this general interest, the number of sequenced mt genomes is increasing rapidly. However, current knowledge is still biased towards higher eukaryotes as exemplified by the number of sequenced chordate mt genomes (1,233) compared to the number of sequenced arthropod mt genomes (271) (METAMIGA-database, September 2009) (Feijao *et al.* 2006). This uneven sampling over the tree of life is also observed on lower taxonomic levels as illustrated when focusing on the Acari (mites and ticks). Although about 90% of all described parasitiform mites belong to the Mesostigmata (Walter and Proctor 1999) only two (METAMIGA-database, September 2009) out of twelve parasitiform mites of which the mt genome is completely sequenced belong to this order, i.e. *Varroa destructor* (Dermanyssina: Varroidae) (Navajas *et al.* 2002) and a predatory mite, *Metaseiulus occidentalis* (Dermanyssina: Phytoseiidae: Typhlodrominae) (Jeyaprakash and Hoy 2007). To obtain a better representation of this order, additional mesostigmatid mt genome sequences are needed. In addition, compared to other metazoan mt genomes, the *M. occidentalis* mt genome shows some unique features. It is unusually large (24,961bp), due to a duplication of a large section of the coding region, and genes coding for NAD6 and NAD3 were not found during the initial annotation process (Jeyaprakash and Hoy 2007). Sequencing the mt genome of a related phytoseiid mite could indicate whether these idiosyncrasies are species specific features. Therefore, we decided to sequence the mt genome of *Phytoseiulus persimilis* Athias-Henriot, 1957 (Acari: Dermanyssina: Phytoseiidae: Amblyseiinae). Phytoseiid mites account for 15% of all mesostigmatid mite species and are the dominant family of predatory mites on plants in most terrestrial ecosystems (Walter and Proctor 1999). In addition, *P. persimilis* was

one of the first natural enemies to be commercialized for spider mite control on a large scale and is still the most widely available species of all marketed invertebrate biocontrol agents (vanLenteren *et al.* 1997) making it an economically important species.

The mt genome of *P. persimilis* was amplified and cloned using a long PCR approach. Size, genome organisation, codon usage and transfer RNA secondary structures were compared with those of the *M. occidentalis* mt genome and a phylogenetic analysis was performed using concatenated mt protein coding genes (PCGs) of complete Acari mt genomes.

2 Materials and methods

2.1 Specimen sampling, DNA extraction and PCR

Specimens of *P. persimilis* and *M. occidentalis* were provided by Biobest (Westerlo, Belgium) and Sterling Insectary (Delano, USA), respectively. Genomic DNA was extracted from approximately 1000 mites using a phenol-chloroform extraction method as described by Van Leeuwen *et al*. (2008). Standard PCR (amplicon < 500bp) was performed in 50 µl volumes (38.5 µl double-distilled water; 5 µl 10 x buffer; 2 mM $MgCl_2$; 0.2 mM dNTP-mix; 0.2 µM of each primer; 1 µl template DNA (± 100ng/µl) and 0.5 µl *Taq* polymerase (Invitrogen, Belgium)). PCR conditions were as follows: 2' 94°C, 35 x (20" 92°C, 45" at a temperature that varies depending on the primers, 1' 72°C) and 2' 72°C. Long PCR (amplicon > 500bp) was performed with the Expand Long Range Kit (Roche, Switzerland) in 50 µl volumes (28.5 µl double-distilled water; 10 µl 5 x buffer; 0.5 mM dNTP-mix; 0.3 µM of each primer; 4 µl 100% DMSO; 1 µl template DNA and 1 µl enzyme-mix). PCR conditions were: 2' 94°C, 10 x (10" 92°C, 20" (exact primers) or 45" (degenerate primers) at a temperature that varies depending on the primers, 1'/kb 58°C), 25 x (10" 92°C, 20" (exact primers) or 45" (degenerate primers) at a temperature that varies depending on the primers, 1'/kb 58°C with 20'' added for every consecutive cycle) and 7' 58°C. All PCR products were separated by electrophoresis on a 0.5 x TAE 1% agarose gel and visualised by EtBr staining. Fragments (amplicon < 1500bp) of interest were excised from gel, purified with the QIAEX II Gel Extraction Kit (Qiagen, Belgium) and cloned into the pGEM-T vector (Promega, Belgium). After heat-shock transformation of *E. coli* DH5α cells, plasmid DNA was obtained by miniprep and inserted fragments were sequenced with SP6 and T7-primers. For each cloned fragment, three independent clones were sequenced. Long PCR products were sequenced by primer-walking on both strands. All sequencing reactions were performed by AGOWA sequencing service (LGC, England).

2.2 Amplification strategy of the mt genome of *P. persimilis*

In an attempt to amplify the complete mt genome of *P. persimilis* in one fragment we designed PCR-primers PP12S-F and PP12S-R (Table III.1), based on a partial *P. persimilis 12S-rRNA* sequence (GenBank accession number: AY099369). This resulted in the amplification of an 8.1kb sequence containing only 6 (*cytB*, *cox3*, *nad2*, *nad3*, *nad4*, *nad4L*)

of the 13 PCG normally present in mt genomes. Hence, an alternative amplification strategy was followed.

Primers LWCOI-F and LWCOI-R (Webster *et al.* 2004) amplified a partial *cox1* sequence from *P. persimilis.* COX1-F (Table III.1), designed from the 3'-end of the partial *cox1* sequence, in combination with the primer COX3-R, designed from the 8.1kb amplicon, successfully amplified a 12kb sequence. Another primer, COX1-R, designed from the 5'-end of the partial *cox1* sequence, in combination with NAD3-F, designed from the 8.1kb sequence amplified a 7.3kb fragment. Assembling the different amplicons resulted in a complete mt genome sequence. The mt genome was finally re-amplified in 3 overlapping pieces with primers PP3a-F, PP3aR, PP3b-F, PP3b-R, PP3c-F and PP3c-R.

Table III.1 - Primers and their sequences used to characterize the *P. persimilis* mt genome and a part of the *M. occidentalis* mt genome.

Primer	Sequence (5' - 3')	T_a (°C)*
PP12S-F	TAAGTGTACATATTGCCCGTCACT	53
PP12S-R	AAGCCGCCAATTAATTTTAGTTTC	53
LWCOI-F	GTTTTGGGATATCTCTCATAC	49
LWCOI-R	GAGCAACAACATAATAAGTATC	49
COX1-F	AGTTGCAAACTTTTGGTTCGTT	58
COX1-R	ATCGCATAAATCATCCCTAACG	58
COX3-R	GGTCTGGAGTAACAGTAACATGAGC	58
NAD3-F	ATTGTGGGGAGGTAAAAGGATT	58
DGNAD3-F	GGRTYRAAHCCRCAYTCAAAIG	49
MO12S-R	GGGTATCTAATCCTAGGTGTTGC	49
MONAD3-R	GAAGAAAGTACGAAGAAGATAAAGAAA	52
MOND4L-F	CCAACCCTAAAATACTATCGCTTAC	52
MOCOX1F	CTTTCTCAACCCGGCTCTTTC	55
MOCOX1R	GTCGAGGTATGCCATTTAGACC	55
LNR1F	TTTGGGTCAAATATTGAAAAGAGA	55
LNR1R	AAATCAACATTCCTCCTATCATTAAAA	55
LNR2F	TGTTAAAAGGTCATTTCGTTTAAAAT	55
LNR2R	GGAAGAATACCTATTGAATTTCCTT	55
LNR4F	GGCTGATAAAAGCGATAAATTG	55
LNR4R	TTTCGCATTTGAATGTACCC	55
PP3a-F	CAGACTTGCCCTGTTAGTCATACC	55
PP3a-R	CAGCATTCTTAGGCTATGTTCTTCC	55
PP3b-F	GGGGTTATTATTGGGTTTGCTG	55
PP3b-R	GGGTTTTTCCCTTTTAGGTCTTG	55
PP3c-F	AATACGGAGCCAAACACTCC	55
PP3c-R	ACCCCTAACCCATCTCATCC	55

* T_a: annealing temperature

2.3 Amplification of the *nad3* gene of *M. occidentalis*

A degenerate primer DGNAD3-F, designed on a conserved region of Acari *nad3*, in combination with the primer MO12S-R, based on the *12S-rRNA* sequence of *M. occidentalis* (GenBank accession number: NC_009093) (Table III.1), amplified a 1.1kb fragment. Another primer, MONAD3-R, designed from the 5'-end of this 1.1kb sequence, in combination with MONAD4L-F, based on the *nad4L* sequence of *M. occidentalis* (GenBank accession number: NC_009093), successfully amplified a 900bp sequence.

2.4 Multipy-primed rolling circle amplification and restriction digests

Extraction and multiply-primed rolling circle amplification (see Fig. II.1 in Chapter II for a scheme of the RCA process) of mtDNA was done according to Van Leeuwen *et al*. (2008). Rolling circle amplified mtDNA of *P. persimilis* was digested with *Hind*III, *Spe*I and *Xba*I (New England Biolabs), cutting the mt sequence at 5, 4 and 3 sites respectively, as predicted from *in silico* analysis. Rolling circle amplified mtDNA of *M. occidentalis* was digested with *Fsp*I and *Spe*I (New England Biolabs), cutting the mt sequence *in silico* at 1 and 4 site(s), respectively. Restriction digest was performed following the manufacturer's instructions and the obtained fragments were subsequently separated by inversed field electrophoresis in 0.5 x TBE 1% agarose gel.

2.5 Southern blotting

MtDNA and rolling circle amplified mtDNA of *M. occidentalis* were both digested with *Fsp*I restriction enzyme (New England Biolabs) and separated by inversed field electrophoresis in 0.5xTBE 1% agarose gel. The enzyme *Fsp*I was chosen as it only cuts once in a non-duplicated region of the *M. occidentalis* mt sequence. After capillary blot, hybridization was performed using the High Prime DIG labeling Starter Kit II (Roche) following manufacturer's instructions with a probe constructed using primers MOCOX1F and MOCOX1R (Table III.1) covering 1,200 bp of the mt genome of *M. occidentalis*. The chemiluminescent signal was visualized on a fluor imager FLA-5100 (Fujifilm, Tokyo, Japan).

2.6 Sequence annotation and analysis

The mt genome sequence of *P. persimilis* was assembled and annotated using VectorNTI 10.0 (Invitrogen, Belgium). Open reading frames (ORFs) were identified with the program getorf from the EMBOSS-package (Rice *et al.* 2000). The obtained ORFs were used as query

in BLASTp (Altschul *et al.* 1997) searches against the non-redundant protein database at NCBI. MatGAT 2.02 was used to calculate similarity and identity values (Campanella *et al.* 2003) between *P. persimilis* and *M. occidentalis* mt proteins. Two large non-protein-coding regions were candidates for the rRNAs (16S and 12S respectively). The boundaries were identified based on alignments and secondary structures of rRNA genes of other mite species. Twenty-two tRNAs were identified by ARWEN1.2.3 (using default settings and the invertebrate mt code) (Laslett and Canback 2008). Secondary structures of tRNAs were drawn in CorelDraw 12.0 (Corel Corporation, Canada). The mt genome sequence of *M. occidentalis* was analysed in a similar way, except that tRNAs were predicted with both ARWEN1.2.3 and tRNA-scan SE (using a cove cut off score of 0.1 and the tRNA-model set to "Mito/Chloroplast") (Lowe and Eddy 1997).

2.7 Phylogenetic analysis

Sequence data were obtained from 22 Acari species and two outgroup taxa, a solifugid (also known as camel spiders) *Nothopuga* sp. and the horseshoe crab *Limulus polyphemus* (for GenBank accession numbers see Table III.2). Only species with a completely sequenced mt genome were selected. Except for *atp8*, *nad4L* and *nad6*, alignments from all mt protein-coding genes (PCGs) were used in phylogenetic analysis. Amino acid sequences and nucleotide sequences were aligned by Clustal W (Thompson *et al.* 1994) as implemented in BioEdit 7.0.1 (Hall 1999). The nucleotide alignment was generated based on the protein alignment using codon alignment. Ambiguously aligned parts were omitted from the analysis by making use of Gblocks 0.91b (Castresana 2000), with default block parameters except for changing "allowed gap positions" to "with half". As Abascal *et al*. (2006) presented evidence that some insects and ticks use a modified mt code (AGG coding for lysine rather than serine), all positions corresponding to AGG codons were removed in the final amino acid alignment.

For the nucleotide alignments the "codons" option was used in Gblocks 0.91b. Due to the results of a saturation analysis on single codon positions (Xia *et al.* 2003), third codon positions were eliminated from the nucleotide alignment. Model selection was done with ProtTest 1.4 (Abascal *et al.* 2005) for amino acid sequences and with Modeltest 3.7 (Posada and Crandall 1998) for nucleotide sequences. According to the Akaike information criterion, the mtART+I+G+F (Cao *et al.* 1994; Abascal *et al.* 2007) model was selected for phylogenetic analysis with amino acid alignments whereas the GTR+I+G model was optimal for analysis with nucleotide alignments. Two different analyses were performed on the amino

acid and the nucleotide dataset. (1) Maximum likelihood (ML) analysis was performed using Treefinder (Jobb *et al.* 2004) whereby bootstrap values were generated from 1000 pseudoreplicates (2) Bayesian inference (BI) was done with MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001). The mtART was implemented in MrBayes according to Rota-Stabelli *et al*. (2009). Four chains ran for 1 000 000 generations, while tree sampling was done every 100 generations. As stationarity was reached after 100 000 generations, the first 1000 trees were used as burnin. The remaining trees were used to calculate Bayesian posterior probabilities.

Table III.2 - Organisms and their GenBank accession numbers used for phylogenetic analysis

Acari species	GenBank accession number
Amblyomma triguttatum	AB113317
Ascoschoengastia sp	AB300501
Carios capensis	AB075953
Dermatophagoides pteronyssinus	EU884425
Haemaphysalis flava	AB075954
Ixodes hexagonus	AF081828
Ixodes holocyclus	AB075955
Ixodes persulcatus	AB073725
Ixodes uriae	AB087746
Leptotrombidium akamushi	AB194045
Leptotrombidium deliense	AB194044
Leptotrombidium pallidum	AB180098
Metaseiulus occidentalis	EF221760
Ornithodoros moubata	AB073679
Ornithodoros porcinus	AB105451
Rhipicephalus sanguineus	AF081829
Steganacarus magnus	EU935607
Tetranychus urticae	EU345430
Unionicola foili	EU856396
Varroa destructor	AJ493124
Walchia hayashii	AB300500
Outgroup species	**GenBank accession number**
Limulus polyphemus	AF216203
Nothopuga sp.	EU024482

3 Results and discussion

3.1 Genome size

The complete mtDNA sequence of *P. persimilis* consists of 16,199bp (GenBank accession number: GQ222414) Fig. III.1, Table III.3). This size was confirmed by restriction digests on rolling circle amplified mtDNA (Fig. III.2*A*) and lays within the range of a canonical metazoan mt genome. Remarkably, the genome is considerably smaller than the mt genome of *M. occidentalis* (24,961bp) which is a member of the related subfamily of the Typhlodrominae (Jeyaprakash and Hoy 2007). To date, the *M. occidentalis* mt genome is one of the largest arthropod mt genomes that have been sequenced. In 2009, at the time this study was carried out, other large arthropod mt genome sequences were found in *Pissodes* sp. (~36kb) (Boyce *et al.* 1989), *Drosophila melanogaster* (19,517bp) (Lewis *et al.* 1995) and *Trialeurodes vaporariorum* (18,414bp) (Thao *et al.* 2004). However, while the size of these genomes are mainly due to repeats of non-coding regions, the large size of the *M. occidentalis* mt genome is the result of an exact duplication of a 9.9kb region containing both PCGs and tRNAs (Jeyaprakash and Hoy 2007). Puzzled by the unexpected discrepancy in size between the genomes of two related species we decided to recheck the size of the *M. occidentialis* mt genome using similar tools as applied for the *P. persimilis* genome. Surprisingly, a restriction digest on rolling circle amplified mtDNA of *M. occidentalis* suggested that the genome is smaller, i.e. approximately 20kb instead of 25kb (Fig. III.2*B1*). To strengthen this finding, we confirmed the size by Southern Blot on enzyme digested mtDNA and rolling circle amplified mtDNA of *M. occidentalis* (Fig. III.2*B2*). Although the difference between our size estimations and previously published data could be the result of strain differences (Fujita *et al.* 2007), it is more likely that the presence of identical control regions in the *M. occidentalis* mt genome could have caused PCR-artefacts during the initial amplification process. Similarly, a first attempt to amplify the complete *P. persimilis* mt genome in one reaction resulted in a 8.1kb fragment. After final annotation of the complete mt genome, it appeared that this initial amplicon was located between the identical large non-coding regions LNR2 and LNR4 (Fig. III.1). Alternatively, small overlaps between PCR fragments (19-345bp, or undefinable due to lack of information) could have hampered the assembling process of the *M. occidentalis* genome (Jeyaprakash and Hoy 2007).

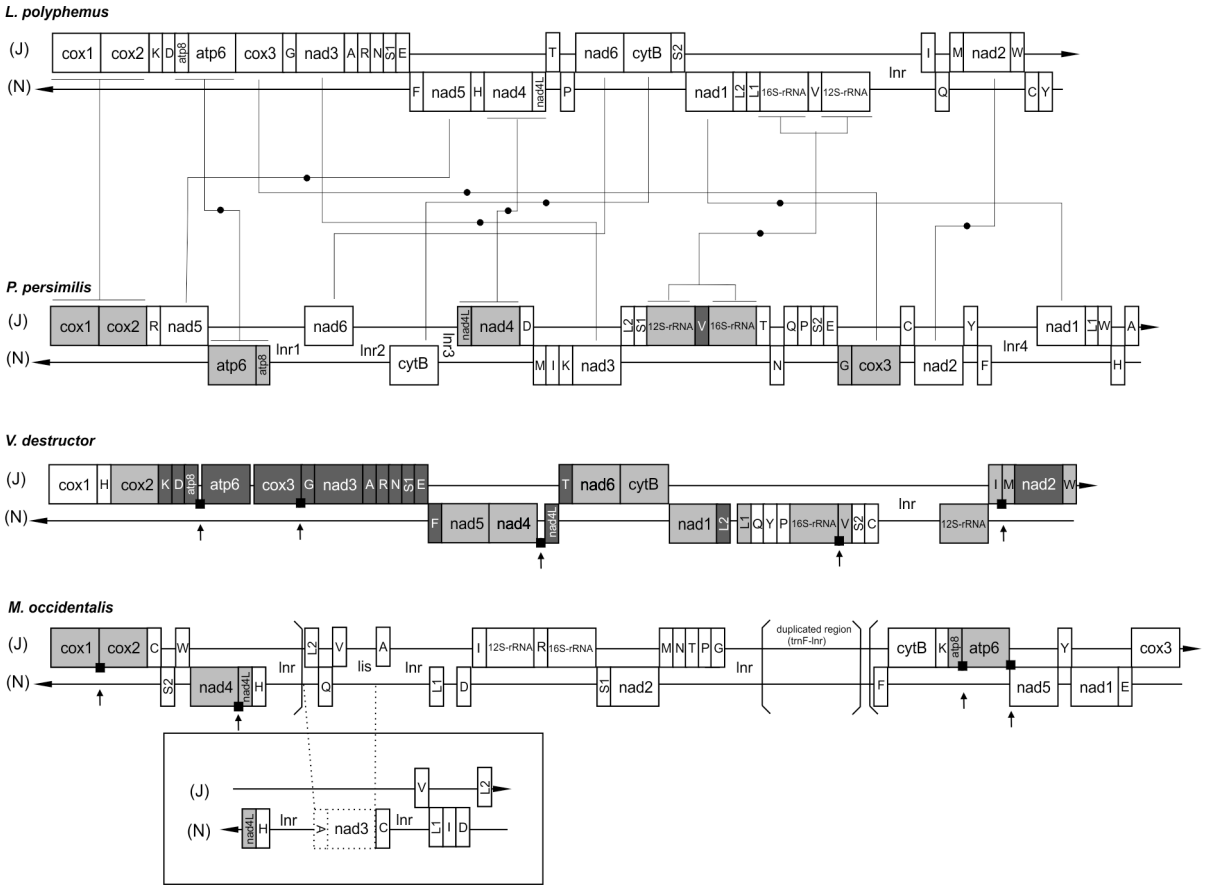


Figure III.1 - Comparison of linearised mt gene orders of *L. polyphemus*, *P. persimilis*, *M. occidentalis* and *V. destructor*.

Graphical linearization of mt genomes is presented according to (Fahrein *et al.* 2007). Gene sizes are not drawn to scale. J stands for majority and N for minority strand. Protein coding and rRNA genes are abbreviated as in the abbreviations section. tRNA genes are abbreviated using the one-letter amino acid code, with L_1=CUN; L_2=UUR; S_1=AGN; S_2=UCN. lnr = plausible control region, lis = large intergenic spacer, other non-coding regions (> 50bp) are indicated as gaps between genes. Braces accentuate the duplicated region in the mt genome of *M. occidentalis* (Jeyaprakash and Hoy 2007). Black circles indicate switches in protein coding gene or rRNA position between J- and N-strands. Dark grey boxes indicate conserved gene boundaries at the 5'- and 3'-end compared to *L. polyphemus*. Light grey boxes indicate conserved gene boundaries at the 5'- or 3'-end compared to *L. polyphemus*. Black squares (indicated with arrows) between the genes of *M. occidentalis* and *V. destructor* represent conserved gene boundaries compared to *P. persimilis*. The location of the cloned *nad3* gene in the mt genome of *M. occidentalis* (this work) is enclosed in a box at the bottom.

Table III.3 - Organisation of the *P. persimilis* mt genome

Gene	Strand[a]	Position	Size (bp)	GC-skew	AT-skew	Start codon	Anti-codon	Stop codon	I[b]
cox1	J	1 - 1,538	1,538	0.176	-0.248	ATG		TA	0
cox2	J	1,539 - 2,207	669	0.283	-0.113	ATG		TAG	5
trnR	J	2,213 - 2,275	63				UCG		33
nad5	J	2,309 - 3,982	1,674	0.268	-0.163	ATA		TAA	27
atp6	N	4,010 - 4,675	666	-0.394	-0.067	ATG		TAA	-7
atp8	N	4,669 - 4,842	174	-0.360	0.087	ATA		TAA	555
nad6	J	5,398 - 5,871	474	0.348	-0.089	ATA		TAA	578
cytB	N	6,450 - 7,538	1,089	-0.183	-0.106	ATA		TAA	112
nad4L	J	7,651 - 7,920	270	0.632	-0.216	ATT		TAA	-11
nad4	J	7,910 - 9,233	1,324	0.380	-0.179	ATA		T	0
trnD	J	9,234 - 9,295	62				GUC		11
trnM	N	9,307 - 9,369	63				CAU		3
trnI	N	9,373 - 9,434	62				GAU		5
trnK	N	9,440 - 9,503	64				CUU		3
nad3	N	9,507 - 9,842	333	-0.246	-0.098	ATA		TAA	12
trnL2	J	9,855 - 9,914	60				UAA		10
trnS1	J	9,925 - 9,978	54				UCU		0
12S-rRNA	J	9,979 - 10,689	711	0.252	-0.007				15
trnV	J	10,705 - 10,764	60				UAC		17
16S-rRNA	J	10,782 - 11,980	1,199	0.337	-0.020				0
trnT	J	11,981 - 12,045	65				UGU		0
trnN	N	12,046 - 12,110	65				GUU		-2
trnQ	J	12,109 - 12,175	67				UUG		-1
trnP	J	12,175 - 12,234	60				UGG		0
trnS2	J	12,235 - 12,299	65				UGA		7
trnE	J	12,307 - 12,367	61				UUC		15
trnG	N	12,383 - 12,443	61				UCC		0
cox3	N	12,444 - 13,227	785	-0.065	-0.074	ATG		T	10
trnC	J	13,238 - 13,291	54				GCA		-2
nad2	N	13,290 - 14,252	963	-0.282	-0.043	ATA		TAA	39
trnY	J	14,292 - 14,351	60				GUA		35
trnF	N	14,387 - 14,455	69				GAA		568
nad1	J	15,024 - 15,938	915	0.354	-0.168	ATT		TAA	5
trnL1	J	15,944 - 16,007	64				UAG		1
trnW	J	16,009 - 16,073	65				UCA		-7
trnH	N	16,067 - 16,127	61				GUG		-4
trnA	J	16,124 - 16,190	67				UGC		9

[a] J, majority strand; N, minority strand
[b] I=intergenic nucleotides and represent the number of nucleotides before the start of the next gene

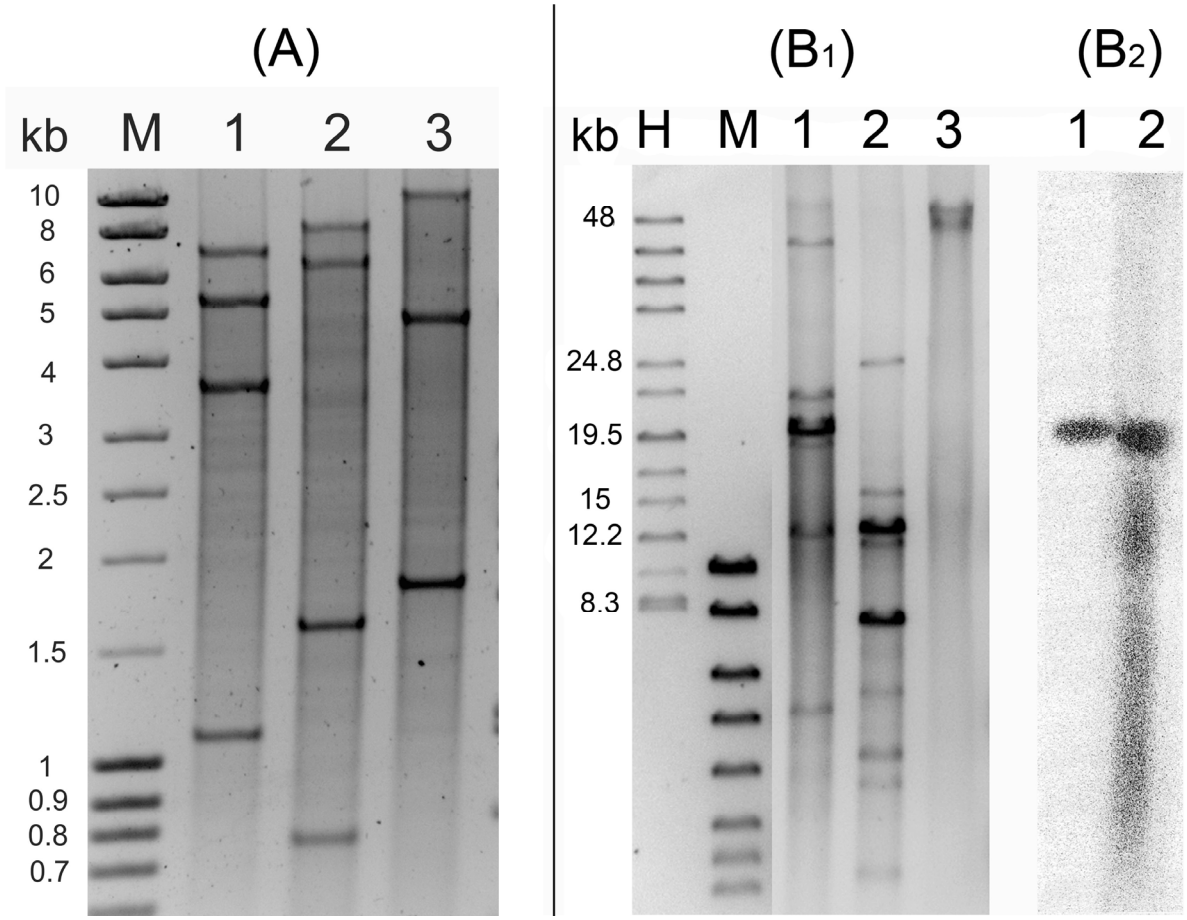


Figure III.2 - Restriction analysis of rolling circle amplified mt DNA of *P. persimilis* (A) and *M. occidentalis* (B).

(A) Rolling circle amplified mtDNA of *P. persimilis* was digested with *Hind*III (1), *Spe*I (2) and *Xba*1 (3); (B1) Rolling circle amplified mtDNA of *M. occidentalis* was undigested (3), digested with *Fsp*I (1) or *Spe*I (2); (B2) Southern blot of purified mtDNA of *M. occidentalis* (1) and rolling circle amplified mtDNA (2) both digested with *Fsp*I; Molecular markers used were MassRuler DNA ladder Mix (Fermentas) (M) and High Molecular Weight DNA Marker (Invitrogen) (H).

3.2 Genome annotation

All 37 genes present in a standard metazoan mt genome could be identified (Fig. III.1, Table III.3) in the *P. persimilis* mt genome. Twenty-four genes are transcribed on the majority strand (J-strand), whereas the others are oriented on the minority strand (N-strand). There are four large (> 50bp) non-coding regions and 20 smaller interspacer regions ranging from 1 to 39bp. Although the mt genome of the horseshoe crab *L. polyphemus* (Chelicerata: Merostomata: Xiphosura) is generally considered to represent the ground pattern for arthropod mt genomes (Staton *et al.* 1997; Lavrov *et al.* 2000), Acari mt gene orders differ markedly from this pattern (see Chapter II) (Domes *et al.* 2008). The *P. persimilis* mt genome forms no exception: only 6 of the 38 gene boundaries in the *L. polyphemus* mt genome are conserved in the *P. persimilis* mt genome (Fig. III.1). Moreover, $trnL_2$ is positioned between *nad3* and $trnS_1$. This is in contrast to an arthropod mt gene order analysis of Boore in which the position of this tRNA between *nad1* and *16S-rRNA* was considered conserved for Chelicerata, Myriapoda and Onychophora (Boore *et al.* 1998). Besides *P. persimilis*, the mt genomes of the Acari *Dermatophagoides pteronyssinus*, *Leptotrombidium* sp., *Steganacarus magnus* and *Tetranychus urticae* also do not share this character state (see Chapter II).

Comparing the *P. persimilis* mt genome with that of two other mesostigmatid mites, *V. destructor* and *M. occidentalis*, revealed that only few gene boundaries are conserved in *P. persimilis* (5 and 4 respectively; Fig. III.1). Remarkably, *P. persimilis* and *M. occidentalis* (both phytoseiid mites) share one gene boundary, *atp6*-3'-end/*nad5*-3'-end, not present in any other arthropod mt genome (Lee *et al.* 2008). Additional mt genomes of other phytoseiid mites should point out if this mt character state is a synapomorphy or rather a homoplasy, since some Cephalopoda (Metazoa: Mollusca) also share this specific gene boundary (Yokobori *et al.* 2004). When the arrangement of tRNAs is ignored, only one extra conserved gene block (*12S-rRNA/16S-rRNA*) could be discovered between *P. persimilis* and *M. occidentalis*, while only two extra conserved gene boundaries (*cox1/cox2* and *12S-rRNA/16S-rRNA*) were present between *P. persimilis* and *V. destructor*. As Cameron *et al*. (2007) also revealed for Phtiraptera mt genomes, the lack of conserved gene blocks between these two phytoseiid mt genomes precludes the accurate reconstruction of the rearrangement events in these genomes.

3.3 Protein coding genes and nucleotide composition

Seven proteins are encoded by genes on the J-strand (*cox1, cox2*, *nad5*, *nad6*, *nad4L*, *nad4*, *nad1*) while six are encoded by genes on the N-strand (*atp6*, *atp8*, *cytB, nad3*, *cox3*, *nad2*). Gene overlap exists between *atp8*/*atp6* (7bp) and *nad4*/*nad4L* (11bp) as is reported for many other mt genomes (Wolstenholme 1992). In contrast to *M. occidentalis*, *P. persimilis* has both *nad3* and *nad6* genes. Wondering whether these closely related mites differ in gene content, we tried to amplify the missing genes of *M. occidentalis*. *Nad3* was amplified in 2 overlapping fragments by PCR (Fig. III.1, Appendix III-A). The gene is located between two duplications of the control region and is transcribed from the N-strand (GenBank accession number: GU066253). The corresponding protein shows 50.5 % identity and 74.8 % similarity with *P. persimilis* NAD3. The amplified region replaces 3 tRNAs (*trnL*$_2$, *trnQ*, and *trnV*) and a large intergenic spacer of the *M. occidentalis* mt genome as annotated by Jeyaprakash and Hoy (2007). In contrast to the ease by which *nad3* was amplified, we were not able to clone fragments of a putative *nad6* sequence, despite the many attempts. It is possible that this gene is indeed lacking as metazoan mt genomes without *nad6* have been reported before (Papetti *et al.* 2007).

Comparing the remaining *P. persimilis* proteins with those of *M. occidentalis*, reveals that cox1, cox2 and cytB are best conserved (90.9%, 80.2% and 80.4% similarity respectively) while atp8 and nad1 showed lowest similarity values (54.4% and 66.8%, respectively) (Fig. III.3). *P. persimilis* PCGs start with either ATG (n=4), ATA (n=7) or ATT (n=2). Ten genes end with TAA or TAG while 3 genes use a truncated stop codon (Table III.3). Polyadenylation of the mRNA is needed in these cases to form a functional TAA stop codon (Ojala *et al.* 1981).

The overall AT-content of the mt genome of *P. persimilis* is 79.84% (Table III.4). This high AT-content is reflected in the codon usage where nucleotides ‘A’ and ‘U’ are preferred over ‘C’ and ‘G’ at the wobble position (see Chapter II). In addition, codons deficient in ‘C’ or ‘G’ are predominantly used. A similar AT-content and codon usage pattern can be observed in the *M. occidentalis* mt genome (Fig. III.4, Table III.4).

In most arthropods there is a strand specific bias in nucleotide frequencies (Hassanin *et al.* 2005; Hassanin 2006). The GC (=%G-%C)/(%G+%C)- and AT (=%A-%T)/(%A+%T)-skew are a good indicator of this strand specific bias (Perna and Kocher 1995). In the *P. persimilis* mt genome the GC- and AT-skew of the J-strand is 0.222 and -0.062, respectively. GC-skews are also positive in all J-strand encoded genes and negative for N-strand encoded genes. The

AT-skews on the other hand are the most negative in the J-strand encoded genes and less negative or slightly positive in the N-strand encoded genes, with *cytB* and *nad3* being exceptions to this trend (Table III.3 and III.4). The calculated skews are reversed to those found in *M. occidentalis* (Table III.4) and most other arthropods and have been reported for only a few species (see Chapter II) (Hassanin *et al.* 2005; Hassanin 2006; Fahrein *et al.* 2007). It has been suggested that such reverse skews are the result of a strand swap of the control region (Hassanin *et al.* 2005) (see Chapter II for a discussion on this topic).

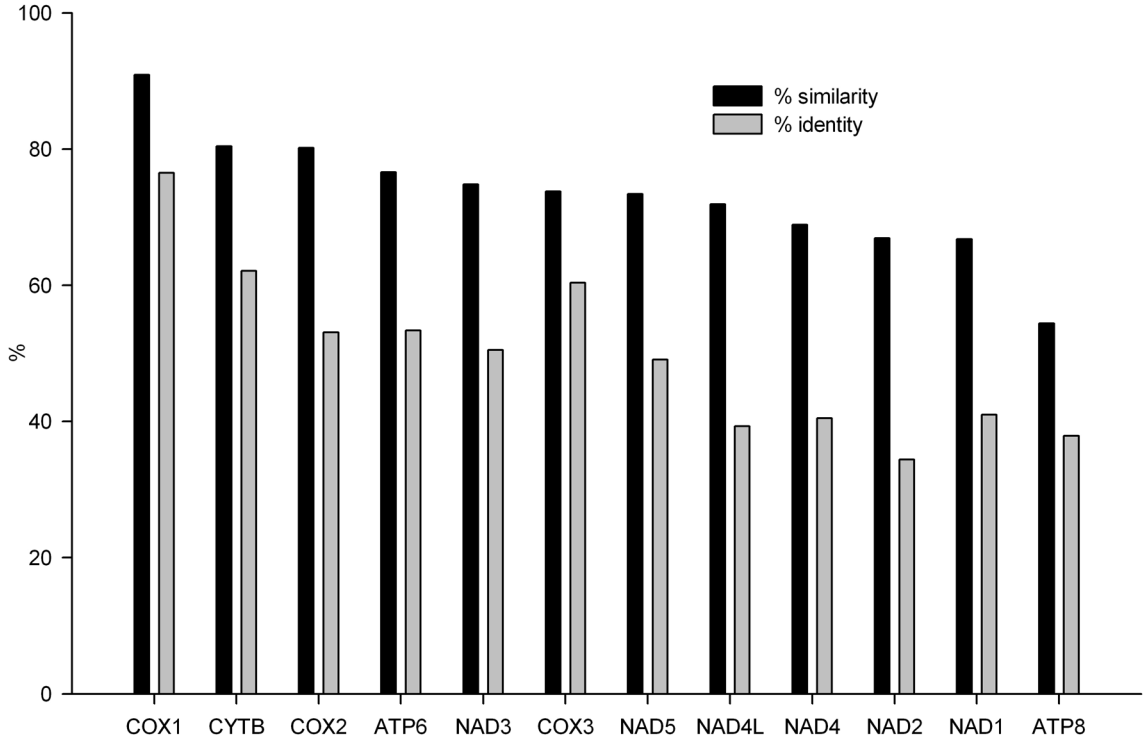


Figure III.3 - Identity and similarity (in %) of mt proteins of *P. persimilis* compared to *M. occidentalis*.

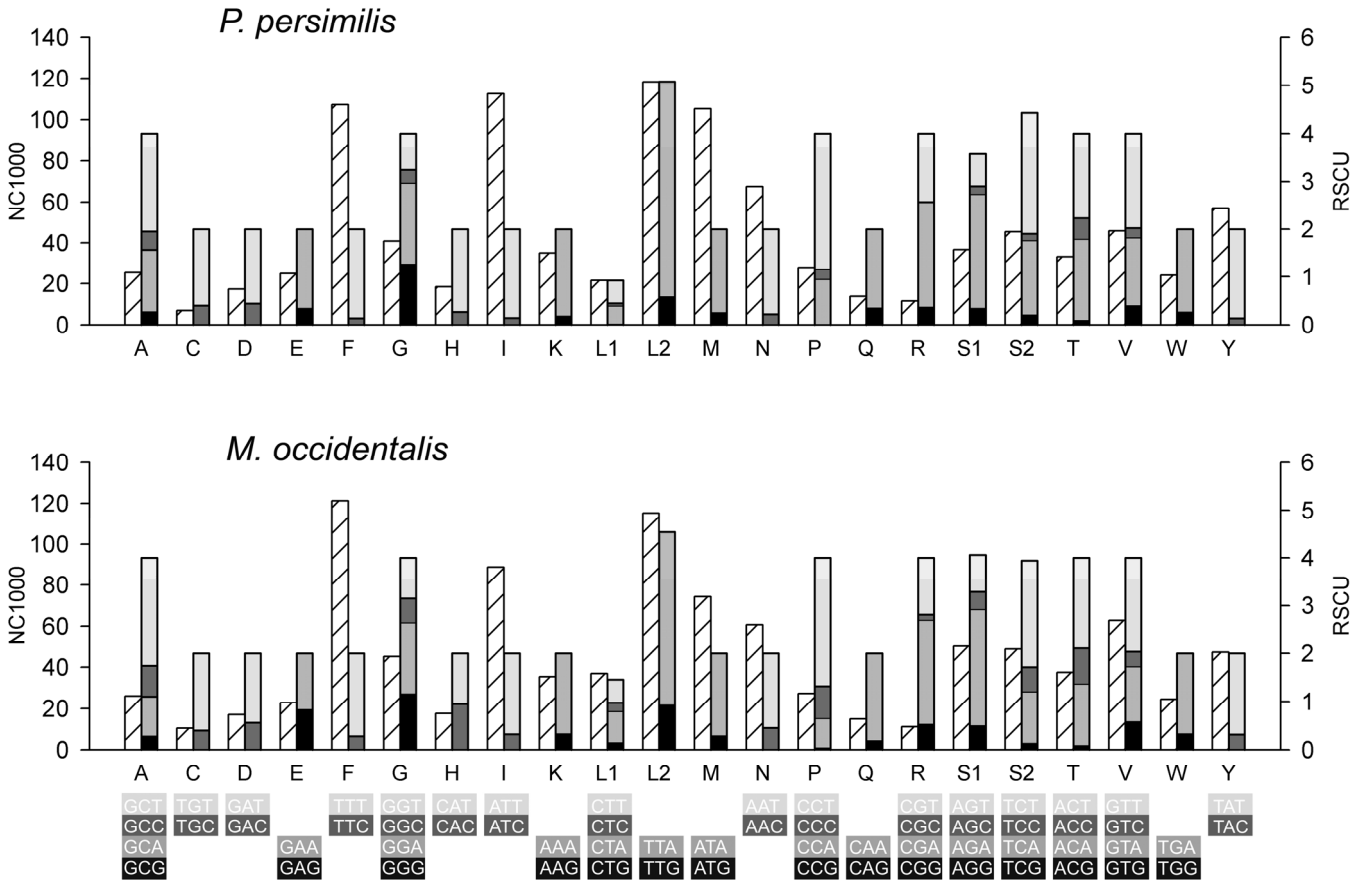


Figure III.4 - Codon distribution and relative synonymous codon usage (RSCU) of mt protein coding genes of *P. persimilis* and *M. occidentalis*.

RSCU is the number of times a particular codon is observed relative to the number of times a codon would be observed in the absence of any codon usage bias (Sharp *et al.* 1988). Grey intensity bars refer to the RSCU-values and have their ordinate on the right. Shaded bars refer to the total number of codons per 1000 codons (NC1000) and have their ordinate on the left. Amino acids and their codons are shown on the x-axis.

Table III.4 - Nucleotide composition of the *P. persimilis* and *M. occidentalis* mt genome[*]

Organism	Complete mt genome				mt PCG [a]		*12S-rRNA*		*16S-rRNA*		Largest non-coding region(s) [b]		References
	Length (bp)	AT%	GC-skew [c]	AT-skew [c]	Length (bp)	AT%	Length (bp)	AT%	Length (bp)	AT%	Length (bp)	AT%	
P. persimilis	16,199	79.84	0.222	-0.062	10,865	79.35	711	81.51	1,199	82.90	112-555-568-578	81.25-75.14-75.09-75.53	This study
M. occidentalis [d]	14,695	76.94	-0.279	0.103	10,014	74.38	742	81.13	1,192	84.31	310-311-311-311	79.35-79.10-79.42-78.78	(Jeyaprakash and Hoy, 2007)

[*] GenBank accession numbers for *P. persimilis* and *M. occidentalis* mt genomes are GQ222414 and EF221760, respectively
[a] mt PCG = mt protein coding genes
[b] duplications of the largest non- coding region were also considered
[c] GC- and AT-skews for the strand coding for the majority of genes, calculated following Perna and Kocher (1995)
[d] length/AT content/GC-skew of the complete mt genome/mt PCGs were obtained from the unique region in the *M. occidentalis* mt genome

3.4 Ribosomal and transfer RNA genes

The AT content of *12S-rRNA* (711bp) and *16S-rRNA* (1199bp) are 82.9% and 81.5% respectively (Table III.3). Both their size and AT content are similar to rRNAs of *M. occidentalis* (Table III.4) or other Acari. The remarkable feature of the rRNAs is their position on the J-strand distant from any of the four largest non-coding regions. In most other arthropods rRNA genes are located on the N-strand in close proximity to the largest non-coding region (control region) (Zhang and Hewitt 1997).

All 22 tRNAs present in a typical metazoan mt genome were traced back in the presented genome: fifteen are encoded on the J-strand and 7 on the N-strand (Table III.3). With the exception of $trnS_1$ (UCU instead of GCU) tRNA anticodon sequences were identical to those most commonly found in other arthropod species. The divergent $trnS_1$ anticodon have also been found in the mt genomes of other parasitiform mites (*Ixodes hexagonus, Rhipicephalus sanguineus,* (Black and Roehrdanz 1998) and *M. occidentalis,* (Jeyaprakash and Hoy 2007)) and an acariform mite (*D. pteronyssinus*, see Chapter II) and is consistent with the codon usage of *P. persimilis*: AGA specifies 89 of the 133 codons encoding Serine-1 while AGC specifies only 6 (Fig. III.4). Secondary structures were predicted for all tRNAs (Fig. III.5). With the exception of *trnK, trnM,* $trnS_l$, all tRNAs have a completely paired acceptor stem (7bp) and a perfect anticodon stem (5bp). *trnM* and $trnS_1$ have a single mismatch whereas *trnK* has two mismatches in the anticodon stem. As mentioned before, sequences of some tRNAs overlap with neighbouring genes (*trnN*/*trnQ* (2bp), *trnQ*/*trnP* (1bp), *trnC/nad2* (2bp), *trnW*/*trnH* (7bp) and *trnH*/*trnA* (4bp)). However, stem mismatches and sequence overlap are not uncommon for mt tRNAs of arachnids (2004; 2008), and are probably repaired by a post-transcriptional editing process (Masta 2000).

Only two tRNAs lacks the D-arm: $trnS_1$ and *trnC*. Whereas the truncated structure of $trnS_1$ is found in most metazoans, the D-arm replacement loop in *trnC* is less common. Nevertheless it has been previously reported for several animal mtDNAs, including two metastriate ticks, *Rhipicephalus sanguineus* and *Boophilus microplus,* and the mesostigmatid *V. destructor* (Black and Roehrdanz 1998; Lavrov *et al.* 2000; Navajas *et al.* 2002). All *P. persimilis* tRNAs possess T-arms. Although similar tRNAs were described for ticks (Black and Roehrdanz 1998) and *V. destructor* (Navajas *et al.* 2002), all tRNAs of the related *M. occidentalis* have T-arm variable (TV)- replacement loops instead of T-arms. This TV-replacement loop is a typical feature for tRNAs of Nematoda and the chelicerate groups of

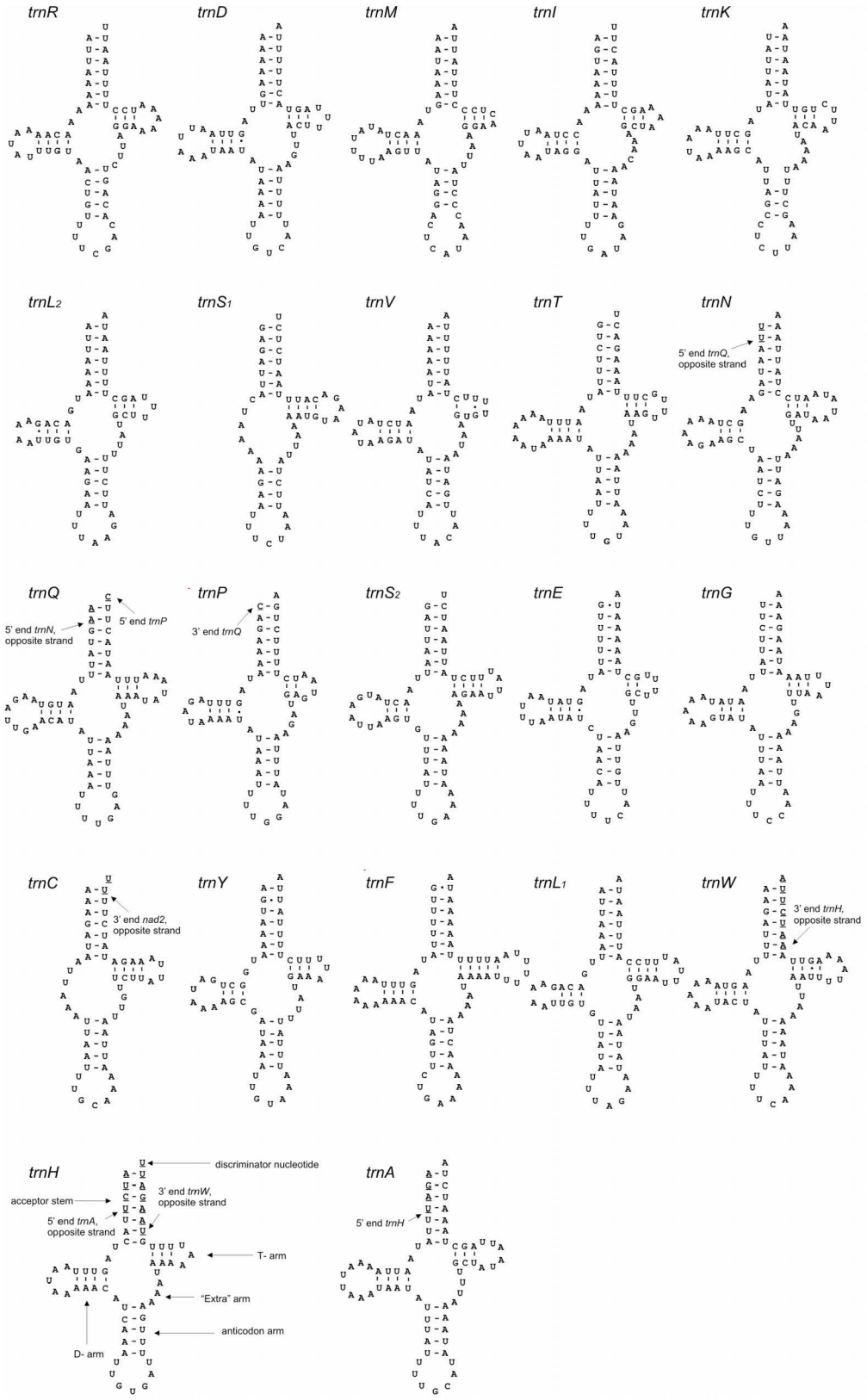


Figure III.5 - Inferred secondary structures of the 22 mt tRNAs from *P. persimilis*

tRNAs are shown in the order of occurrence in the mt genome starting from *trnR*. Locations of adjacent gene boundaries are indicated with arrows. Underlined font indicates that the sequence is part of the adjacent gene. Inferred Watson-Crick bonds are illustrated by lines, whereas GU bonds are illustrated by dots.

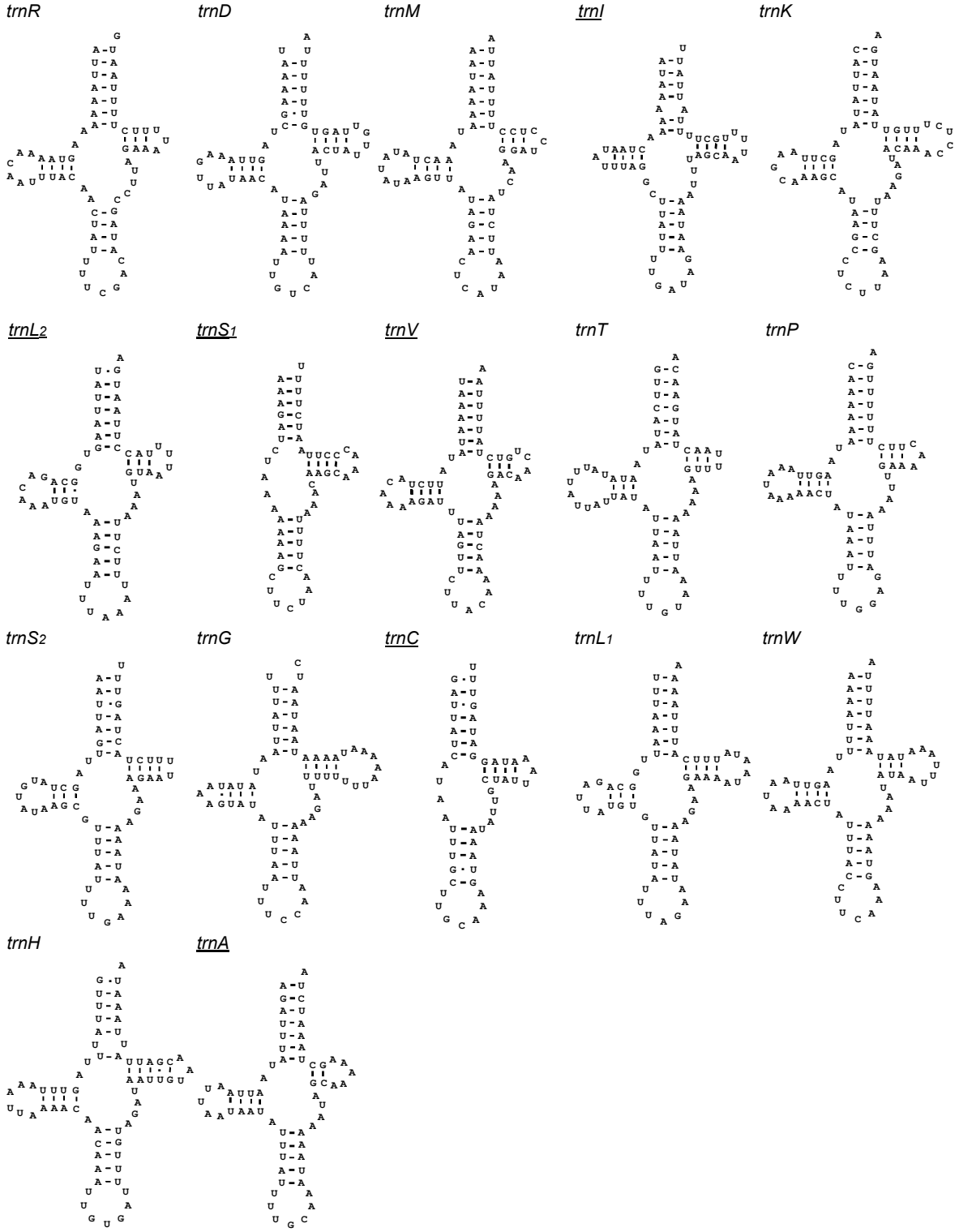


Figure III.6 – Alternative secondary structures for 17 tRNAs of the mt genome of *M. occidentalis.*

Underlined tRNAs are located at alternative positions in the mt genome of *M. occidentalis* as annotated by Jeyaprakash and Hoy (2007). Inferred Watson-Crick bonds are illustrated by lines, whereas GU bonds are illustrated by dots

Araneae, Scorpiones, Thelyphonida and Acariformes (see Chapter II) (Masta and Boore 2008; Ernsting *et al.* 2009). However, their appearance in *M. occidentalis* tRNAs could be explained by the fact that they were identified by tRNA-scan SE with the source set to "nematode mito", limiting the search to tRNAs with TV-replacement loops (Jeyaprakash and Hoy 2007). Repeating the analysis using "mito/chloroplast" settings and with the newly amplified *nad3* region integrated in the *M. occidentalis* mt genome, revealed how ten of the predicted *M. occidentalis* mt tRNAs (*trnD*, *trnG*, *trnH*, *trnK*, $trnL_1$, *trnM*, *trnP*, $trnS_2$, *trnT* and *trnW*) hold perfect T-arms instead of TV-replacement loops (Fig. III.6). ARWEN1.2.3 confirmed these findings and found an additional cloverleaf tRNA for *trnR*. In addition, a more conventional structure for $trnS_1$ (loss of D-arm, presence of T-arm) was found at the position where an aberrant *trnC* with disturbed anticodon loop was previously predicted. Finally, ARWEN1.2.3 also predicted cloverleaf secondary structures for *trnA*, *trnI*, $trnL_2$ and *trnV*, at alternative positions in the mt genome and a D-armless *trnC* at the position where *trnA* was originally located (Fig. III.1, Fig. III.6, Appendix III-A) (GenBank accession number: GU066253). All suggested tRNAs aligned better with corresponding tRNAs of other Acari (including *P. persimilis*) than their T-armless versions (data not shown). In addition, the only 2 tRNAs with a T-arm but without D-arm (*trnC* and *trnS1*) are now conserved in the mt genome of *P. persimilis* and *M. occidentalis.* This reannotation of the canonical cloverleaf tRNAs in *M occidentalis* is important as the secondary structures of the novel tRNAs are again in line with the hypothesis that T-arm less tRNAs are mainly preserved in acariform mt genomes whereas tRNAs of parasitiform mt genomes generally have canonical cloverleaf secondary structures (Masta and Boore 2008).

3.5 Large non-coding regions

Four large (> 50bp) non-coding regions (LNR1-4) could be identified in the mt genome of *P. persimilis* (Fig. III.1). LNR1, LNR2 and LNR4 are located between *atp8/nad6* (555bp), *atp6*/*cytB* (578bp) and *trnF*/*nad1* (568bp), respectively (Table III.3, Fig. III.1). They have the same orientation and share a nearly identical 525bp region with a T63A transition in LNR1 as only difference. The position of the three LNRs was confirmed by PCR with primers LNR1F/LNR1R, LNR2F/LNR2R and LNR4F/LNR4R (Table III.3). The PCRs generated amplicons of expected sizes: 944bp, 648bp and 843bp, respectively (Fig. III.7). The other non-coding region (LNR3) is located between *cytB* and *nad4L*. With a length of 112bp it is considerable smaller than the other LNRs. Although this region lacks overall sequence similarity to the other LNRs, a short motif is conserved over the 4 LNRs (5'-AGTGAGA-3').

The mt genome control region has only been experimentally mapped for a small number of arthropods (Cameron *et al.* 2007), but no chelicerates have been examined. Actually, most arthropod mt control regions were allocated solely based on their similarity with control regions of insect mt genomes. A high AT-content, the close proximity of *12S-rRNA*, the potential to form stem-loop structures and the presence of poly-A/T stretches are all common features of insect mt control regions (Zhang and Hewitt 1997; Saito *et al.* 2005). All LNRs can form stem-loop structures and have poly-A/T stretches (data not shown), with the overall AT-content of LNR3 being the highest (81.25% compared to ±75% for the other LNRs). Nevertheless LNR1,2 and 4 have a 100bp stretch with a 92% AT-content (data not shown). However, as none of the LNRs were in close proximity of *12S-rRNA* and stem-loop structures of other Acari mt control regions show a great variability in sequence and size (see Chapter II) (Lavrov *et al*. 2000; Shao *et al*. 2005a; Jeyaprakash and Hoy 2007; Domes *et al*. 2008) it was not possible to determine which of the LNRs is functioning as control region.

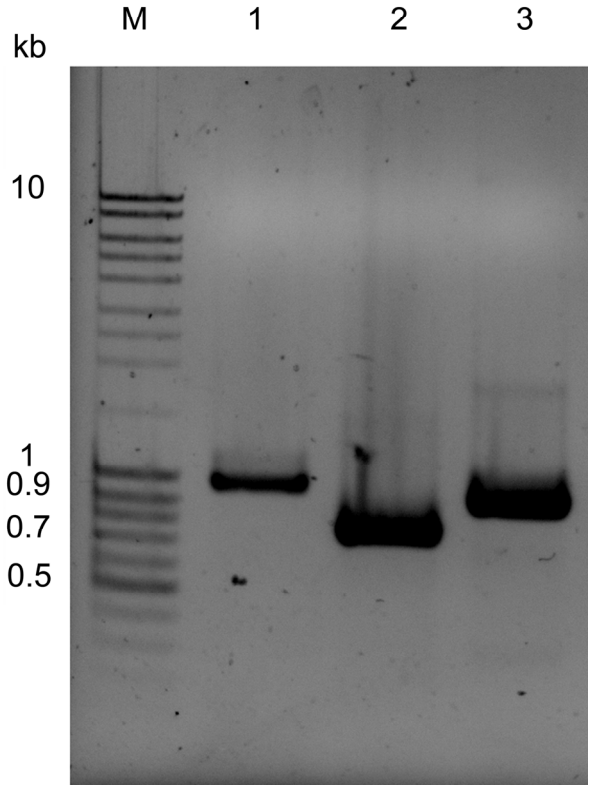


Figure III.7 - PCR amplification of large non-coding regions LNR1, 2 and 4 of the mt genome of *P. persimilis*

(1) LNR1 amplicon (2) LNR2 amplicon and (3) LNR4 amplicon. M: MassRuler DNA ladder Mix (Fermentas).

As multiple LNRs with identical sequences can function as a hot-spot for slipped-strand mispairing (Levinson and Gutman 1987), causing tandem duplication of the spacing region between these regions, LNR1,2 and 4 could explain some of the numerous rearrangements in the *P. persimilis* mt genome. Random deletion of duplicated genes following this duplication would then result in a new mt gene arrangement between these regions (Boore 2000; Kumazawa and Endo 2004). However, other metazoan mt genomes with duplicated large non-coding regions show only slight or no gene rearrangements at all (Shao *et al*. 2005b) Moreover, tandem/duplication followed by deletion can only explain mt rearrangements on the same strand. As several mt genes of *P. persimilis* also changed orientation compared to the ancestral ground pattern (Fig. III.1), other mechanisms like intra-mtDNA recombination (Lunt and Hyman 1997) must have contributed to its current mt gene order.

3.6 Phylogenetic relationships

A phylogenetic tree was constructed based on nucleotide and amino acid sequences from all except *atp8*, *nad4L* and *nad6*, mt PCGs of Acari. ML and BI analysis performed on the nucleotide data set resolved trees with an identical topology in which *P. persimilis* clusters with *M. occidentalis* and forms, together with *V. destructor*, a sistergroup of the Ixodida within the Parasitiformes (Fig. III.8). This is in agreement with the most recent views on the classification of the Parasitiformes (Klompen *et al.* 2007; Dunlop and Alberti 2008). The amino acid data set resulted in similar trees, confirming the evolutionary position of *P. persimilis*. The only major inconsistency over the trees was the position of *T. urticae.* In the ML tree based on the amino acid dataset *T. urticae* formed a sistergroup of the Trombiculidae and *Unionicola foili*, while in the BI tree *T. urticae* clustered with the Trombiculidae, forming a sistergroup of *U. foili*. The latter is in contrast with views on the classification of the Trombidiformes (Lindquist 1996). However, as the position of *T. urticae* in both ML trees is supported by low bootstrap values, additional mt genome data from closely related taxa of *T. urticae* and from taxa located between *U. foili* and *T. urticae* would increase the resolution of this branch.

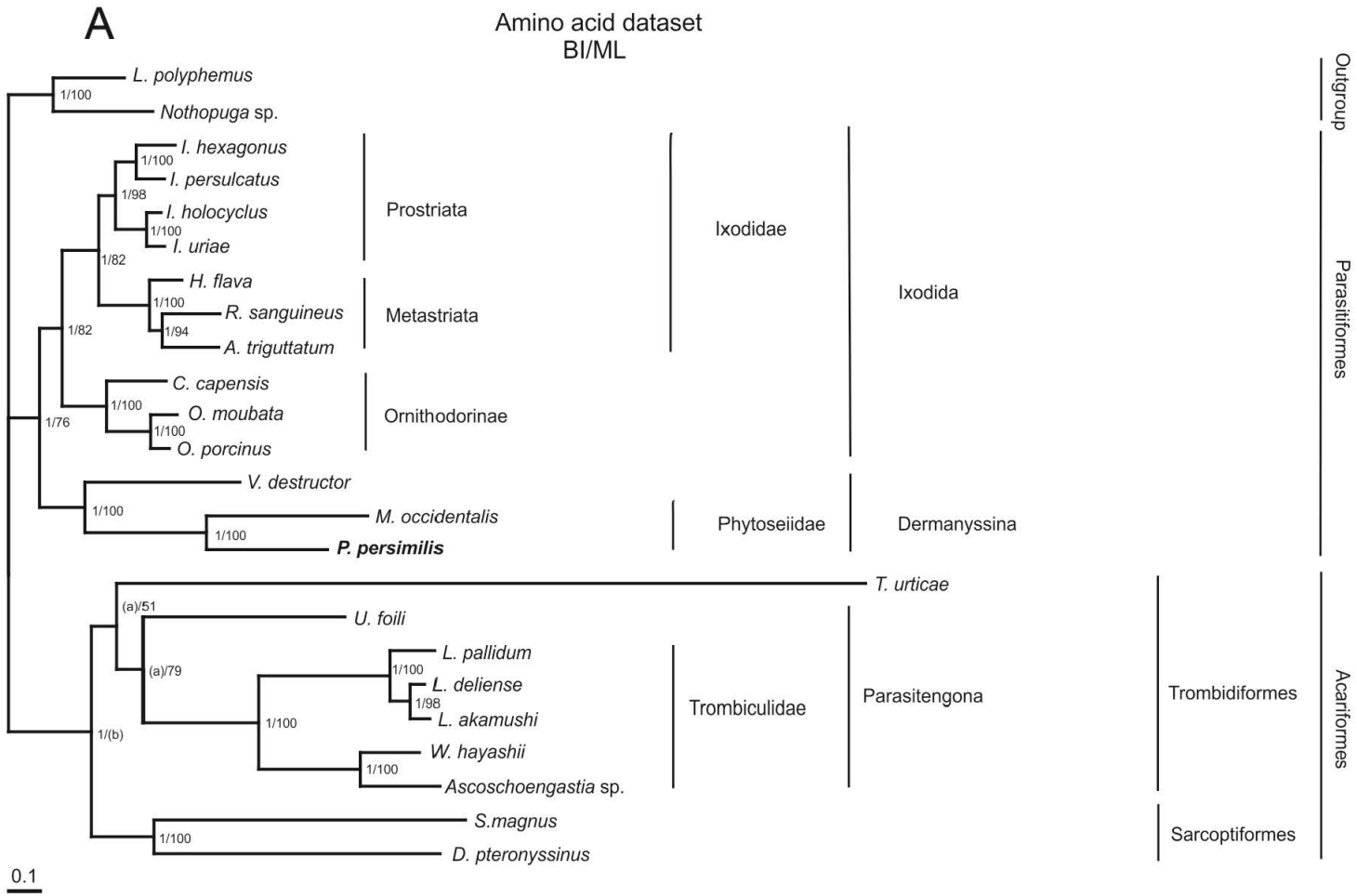

A
Amino acid dataset
BI/ML
L. polyphemus
Nothopuga sp.
I. hexagonus
I. persulcatus
I. holocyclus
I. uriae
H. flava
R. sanguineus
A. triguttatum
C. capensis
O. moubata
O. porcinus
V. destructor
M. occidentalis
P. persimilis
T. urticae
U. foili
L. pallidum
L. deliense
L. akamushi
W. hayashii
Ascoschoengastia sp.
S.magnus
D. pteronyssinus
1/100
1/98
1/82
1/94
1/76
(a)/51
(a)/79
1/(b)
0.1
Prostriata
Metastriata
Ornithodorinae
Ixodidae
Phytoseiidae
Trombiculidae
Ixodida
Dermanyssina
Parasitengona
Trombidiformes
Sarcoptiformes
Outgroup
Parasitiformes
Acariformes

Figure III.8

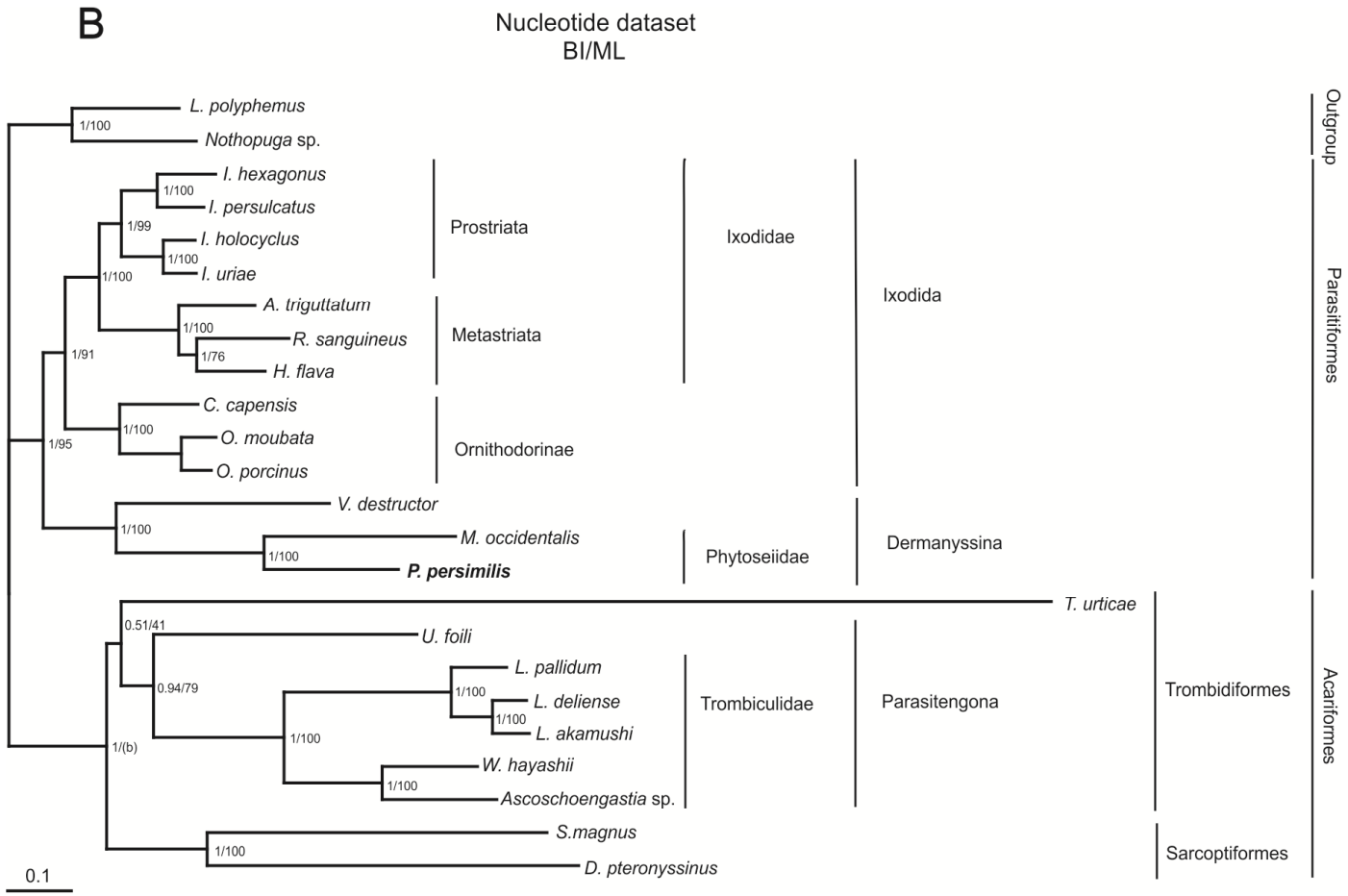


Figure III.8 (continued) - Phylogenetic trees of Acari relationships.

Trees were inferred from amino acid (A) and nucleotide (B) datasets. Trees were rooted with two outgroup taxa (*L. polyphemus* and *Nothopuga* sp.). Numbers at the branch points are percentages from Bayesian posterior probabilities (left) and ML bootstrapping (right). (a) in the ML tree based on the amino acid dataset *T. urticae* forms a sistergroup of the Trombiculidae and *U. foili*, while in the BI tree *T. urticae* clustered with the Trombiculidae, forming a sistergroup of *U. foili*; (b) the relationship between the Sarcoptiformes and the Trombidiformes was not resolved in the ML-trees. Accession numbers for the different Acari and outgroup taxa mt genomes are listed in Table III.2

Chapter IV

A link between host plant adaptation and pesticide resistance in the polyphagous spider mite *Tetranychus urticae*

This chapter has been redrafted from:

Dermauw W., Wybouw N., Rombauts S., Menten B., Vontas J., Grbic M., Clark R.M., Feyereisen R., & Van Leeuwen T. (2013) A link between host plant adaptation and pesticide resistance in the polyphagous spider mite *Tetranychus urticae*. *Proceedings of the National Academy of Sciences of the United States of America* 110 (2): E113-E122

1 Introduction

Plants produce a wide variety of allelochemicals including a plethora of defense compounds. These can affect herbivore fitness in subtle ways by changing behavior or in less subtle ways by causing acute toxicity. The effectiveness of plant defenses is remarkable as herbivory has evolved successfully in only about one third of all animals (Strong *et al.* 1984). Nevertheless, herbivores are among the most diverse terrestrial faunas (Futuyma and Agrawal 2009). The ability to metabolize and detoxify plant chemicals is considered one of the major responses that arthropod herbivores have evolved during their coevolution with plants. Thus, the vast majority of insect herbivores are associated with no more than one or a few plant species (Bernays and Chapman 1994), potentially reflecting the need for specialized mechanisms to cope with plant chemicals. Herbivorous specialists encounter high levels of predictable toxicants and have often evolved efficient and specialized detoxification systems (Cornell and Hawkins 2003). A well-known example is the role of CYP6B enzymes in *Papilio* species that feed on plants containing toxic furanocoumarins (Berenbaum *et al.* 1996). These enzymes, belonging to the large P450 family, can convert these compounds to non-toxic metabolites and are thought to be a key innovation allowing the "escape and radiate" diversification of Papilionidae (Thompson 1994). Plants too can escape and radiate by producing new chemicals that are toxic to herbivores that have not yet evolved an effective detoxification response. An extension of this reasoning is that compounds that have evolved earlier and that are taxonomically widespread should be less toxic than newer compounds, and that specialist herbivores should be less affected than generalists by the toxic compounds of their host plant (Cornell and Hawkins 2003). This is the "jack of all trades, master of none" argument comparing generalist/specialist ability to cope with plant secondary chemistry (Ali and Agrawal 2012). However, the way generalist (polyphagous) herbivores cope with the tremendous variety of chemicals in their toxic diet is not well documented in molecular terms. The original assumption was that generalists have a greater capacity to detoxify plant chemicals than specialists (Krieger *et al.* 1971). This has been refined to state that generalists have detoxification enzymes, in particular P450 enzymes, with broader substrate specificity (Berenbaum *et al.* 1992).

Recently introduced chemical pesticides can be considered as a metaphor for newly evolved or encountered plant chemicals, and a parallel has often been drawn between the evolution of resistance to insecticides and the response to host plant chemicals. This view was presented

by Gordon in 1961, who thought that resistance genes are alleles of common genes, "the normal function of which is metabolism of biochemicals present in the diet" (Gordon 1961). The "pre-adaptation hypothesis" for insecticide resistance has been supported by surveys of the literature (Croft and Strickler 1983; Rosenheim *et al.* 1996) although the comparisons drawn between herbivores and natural enemies, or between chewing and sucking herbivores may be confounded by taxonomy, thus calling for other forms of experimental and observational evidence (Rosenheim *et al.* 1996). It is now well accepted that herbivore exposure to different plant allelochemicals can affect the toxicity of pesticides (Brattsten *et al.* 1977; Yu *et al.* 1979; Berry *et al.* 1980; Gould *et al.* 1982; Kennedy 1984; Ahmad 1986; Lindroth 1989; Li *et al.* 2000; Yang *et al.* 2001; Li *et al.* 2004; Sasabe *et al.* 2004; Sen Zeng *et al.* 2007; Castle *et al.* 2009). Moreover, metabolic resistance to pesticides is known to commonly rely on the increased expression of one or more genes encoding detoxification enzymes and formal evidence that many of these detoxification enzymes can metabolize both plant chemicals and pesticides is accumulating (Li *et al.* 2007; Feyereisen 2012). However, it has also been argued that the pattern of selection by plant allelochemicals and by pesticides differs (Despres *et al.* 2007; Li *et al.* 2007), so whether the polyphagous nature of many crop pests results in a pre-adaptation potential to cope with pesticides remains conjectural.

To elucidate the relationship between host plant adaptation and pesticide resistance in a systematic way, the two-spotted spider mite, *Tetranychus urticae*, is an excellent choice. *T. urticae* is among the most polyphagous herbivores known: it can feed on over 1,100 different plants in more than 140 different plant families that produce a broad spectrum of chemical defenses (Grbić *et al.* 2011). Spider mites have been shown to rapidly adapt to new or less favorable hosts without a correlated fitness cost when compared to the ancestral host (Agrawal 2000; Magalhães *et al.* 2009). Moreover, long-term adaptation on a single host does not markedly reduce genetic variation, or the capability to subsequently adapt to a different host (Fry 1989; Magalhães *et al.* 2007). Also, experimental evolution has shown that although induced plant responses to *T. urticae* herbivory decrease the fitness of unadapted mites, induced plant response resulted in higher fitness of adapted mites, suggesting that spider mites can overcome both constitutive and induced plant defenses (Agrawal 2000). In parallel with an exceptionally broad host range, *T. urticae* has demonstrated an unprecedented ability to develop resistance to pesticides; regardless of the chemical class, the first cases of resistance are usually reported within a few years after the introduction of a new acaricide. Selection for resistance in *T. urticae* is accelerated by its high fecundity and very

short life cycle (Van Leeuwen *et al*. 2010a), and potentially also its haplodiploid sex-determination system (unmated females produce haploid males) (Denholm *et al.* 1998; Carriere 2003).

To date, studies of resistance in *T. urticae* have focused largely on target site mutations and on classical detoxifying enzyme systems, such as P450 monooxygenases (P450s), carboxyl/cholinesterases (CCEs) and glutathione-S-transferases (GSTs) (Li *et al*. 2007; Van Leeuwen *et al*. 2010a). However, these studies have not been satisfactory for understanding the scope of acaricide resistance in *T. urticae*. Under field conditions, multi-resistant strains that are resistant to all commercially available acaricides are often encountered, and strikingly these strains also resist compounds with new modes of action that have never been used in the field (Khajehali *et al.* 2011). In this study, we have taken advantage of the high quality genome sequence of *T. urticae* (Grbić *et al*. 2011; Van Leeuwen *et al*. 2013) to construct an expression microarray that we then used to collect genome-wide expression data over a time course ranging from hours to generations after transfer of mites to a new, challenging host. We then related changes in gene expression after host plant change to constitutive patterns of gene expression in two strains that are highly resistant to a spectrum of pesticides. In doing so, we defined a set of genes and gene families that are of potential adaptive relevance to both situations. Remarkably, our studies suggest that the polyphagous spider mite exploits a large and shared repertoire of "classical" detoxification genes as well as potential new players as a defense against plant chemicals and pesticides.

2 Materials and Methods

2.1 Plant rearing

Tomato-seeds (*S. lycopersicum* cv. 'Moneymaker') and kidney bean (*P. vulgaris* cv. 'Prelude') were potted in black earth (Structural Professional, pH 5.0-6.5, 20% organic matter (Snebbout NV, Belgium) and allowed to grow in a growth chamber at 26+0.5°C, 60% relative humidity (RH) and an 16:8 h Light:Dark (L:D) photoperiod. Tomato plants were used for experiments when they had at least four completely developed leaves (about 35 days old) while bean plants were used for either experiments or spider mite rearing when they had 2 completely developed leaves (about 14 days old).

2.2 Mite strains

The London reference strain originates from a wild-collected *T. urticae* population from the Vineland region (Ontario, Canada) and DNA from an inbred line of this strain was used for *T. urticae* genome sequencing (Grbić *et al.* 2011). This strain is susceptible to commercially available acaricides (Khajehali *et al.* 2011). The LS-VL laboratory reference strain, originally collected in 2000 near Ghent (Belgium), has been previously described as highly susceptible to acaricides (Van Leeuwen *et al.* 2005). The MR-VP resistant strain was originally collected from different cultivars of bean plants in a greenhouse at the national botanical garden (Brussels, Belgium) in September 2005, where spider mite control was reported to be extremely problematic. The strain was controlled by regular foliar applications of commercial formulations of the following acaricides: tebufenpyrad, pyridaben, clofentezine, hexythiazox, bifenthrin, fenbutatin oxide, abamectine and oxamyl. Resistance to Mitochondrial Electron Transport Inhibitor (METI) acaricides is well characterised in MR-VP (Van Pottelberge *et al.* 2009). The Marathonas (MAR-AB) strain was isolated from a heavily sprayed rose greenhouse near Athens (Greece) in 2009. The strain is highly resistant to abamectin, bifenthrin, clofentezine, hexythiazox, fenbutatin oxide and pyridaben.

All *T. urticae* strains were mass reared on potted kidney bean plants in a climatically controlled room at 26°C (±0.5°C), 60% RH and 16:8 h L:D photoperiod. The strains were offered fresh bean plants weekly.

2.3 Host change experiment

For each timepoint (2h, 12h, 80 days) a tomato plant was infested with about hundred fifty female mites from the London strain (cultured on bean plants). Two (Tomato-2h), 12 hours

(Tomato-12h) and 80 days (about 5 generations, Tomato-5G) after infestation 100 mites were re-collected for total RNA extraction. All experiments were performed at 26+0.5°C, 60% RH and an 16:8 h L:D photoperiod and four biological replicates were performed for each time point. The Tomato-2h and Tomato-12h experiments were performed during the 16h light photoperiod.

2.4 Toxicity tests

Toxicity bioassays on the London strain with and without adaptation to tomato (8 generations) were performed in a similar way as described by Van Leeuwen *et al*. (2005). First, adult female mites were transferred to square kidney bean leaf discs (*P. vulgaris* cv. 'Prelude'), placed on wet cotton in a Petri dish. Subsequently, 800 µl of the LC_{90} (lethal concentration killing 90% of the population) of acaricides with different modes of action (pyridaben (Sanmite 200 mg/g WP), tebufenpyrad (Pyranica 200 g/L EC): mitochondrial complex I electron transport inhibitors; milbemectin (Milbeknock 9.3 g/L EC): chloride channel activator; fenbutatin oxide (Torque 550 g/L SC): inhibitor of mitochondrial ATP synthase and bifenthrin (Talstar 80 g/L SC): sodium channel modulator) was sprayed on the mites. Mites sprayed with double distilled water were used as a control. Finally, bioassays were placed in a climatically controlled room at 26±0.5°C, 60% RH and 16:8 h L:D photoperiod. Four replicates were performed for each strain and for each acaricide. Mortality was assessed 24 hours after acaricide application and corrected for control mortality using Abott's formula (Abbott 1925). Mites were considered dead when they were drowned or when they did not move after prodding with a fine hair paintbrush.

2.5 Microarray experiments and qPCR

2.5.1 Microarray construction

A custom Sureprint genome wide G3 Gene Expression 8x60K microarray was designed using the Agilent eArray platform (Agilent Technologies) based on the *T. urticae* genome annotation (version from April 20, 2010). The probe design aimed for 3 probes of 60 nucleotides per gene with a T_m of 80°C and parameters set to "best probe design" and "3' bias". In order to also design gene-specific probes for highly similar genes (e.g. duplicated genes), coding sequences were extended with 100 bp of their predicted 3' UTR. Where 3' UTRs were not predicted or predicted shorter, 100 bp downstream the stop-codon were added to the coding sequence. In total 58,985 probes were designed. Before the start of GeneSpring analysis (see below), probes were re-mapped (using Bowtie, version 0.12.7 (Langmead *et al.*

2009)) on the most recent genome annotation (April 18, 2011, 18455 predicted genes). According to this mapping, 87.4% from the latter were covered by at least 1 probe, while 81.7 % were covered by at least 3 probes. Standard Agilent features such as spike-ins were added (IS-62976-8-V2_60kby8_GX_EQC_20100210). We selected 182 unique probes that mapped to *T. urticae* genes expressed across 4 developmental stages, as identified by RNAseq experiments (Grbić *et al.* 2011), with different ranges of expression (based on normalised read-counts (RPKM)). These probes (probe names starting with "Rep") were randomly distributed in 10 to 15 copies per array, and were used to measure array reproducibility. The array design was submitted to NCBI under the GEO-platform format (GPL15756).

2.5.2 Target preparation, microarray hybridisation and analysis

Total RNA was extracted from 100 1-3 day old adult female mites using RNeasy mini kit (Qiagen) in 4-6 replicates for each strain (MR-VP, MAR-AB, London) or for each time point (Tomato-2h, Tomato-12h, Tomato-5G) in the host change experiment. Contaminating DNA was removed by digestion with RNase-free Turbo DNase (Ambion). The quality and quantity of the RNA were assessed by NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies) and by running an aliquot on a 1% agarose gel. One hundred nanograms of RNA was used to generate Cy3- and Cy5-labeled cRNA, using the Agilent Low Input Quick Amp Labeling Kit (version 6.5, Agilent Technologies). RNA spike-in controls (Agilent Technologies) were added to RNA samples before cRNA synthesis. The labeled cRNA was purified with the RNeasy mini kit (Qiagen), dye content (> 8.0 pmol dye/µg cRNA) and the concentration of cRNA was measured by NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies). Cy3- and Cy5-labeled cRNAs were pooled and hybridized using the Gene Expression Hybridization Kit (Agilent Technologies) for 17 h in a rotating hybridization oven at 20 rpm and 65°C. Following hybridisation experiments were performed (the number of biological replicates is given between brackets): Cy5 labeled MR-VP cRNA/Cy3 labeled London cRNA (6), Cy5 labeled MAR-AB cRNA/Cy3 London cRNA (5), Cy5 labeled Tomato-2h cRNA/ Cy3 labeled London cRNA (4), Cy5 labeled Tomato-12h cRNA/ Cy3 labeled London cRNA (4), Cy5 labeled Tomato-5G cRNA/ Cy3 labeled London cRNA (4). After hybridization, slides were washed using the Gene Expression Wash Buffer kit (Agilent Technologies), treated with Stabilization and Drying solution (Agilent Technologies), protected by an Ozone-Barrier cover (Agilent Technologies) until scanned by an Agilent Microarray High Resolution Scanner with default settings for 8 x 60K G3 microarrays. Data

were then normalized by the Agilent Feature Extraction software version 10.5 (Agilent Technologies) with default parameter settings for gene expression two-color microarrays (protocol GE2_105_Dec08) and data was transferred to GeneSpring GX 11.0 software (Agilent Technologies) for further statistical evaluation. Experiments were constructed from these microarray data with GeneSpring GX 11.0. Next, probes were flag filtered (only probes that had flag-value 'Present' in 50% of all replicates of each experiment were retained) and linked to the most recent annotation file (April 18, 2011) using the 'Create New Gene-Level Experiment"-option. Genes with a Benjamini-Hochberg FDR value < 5% and having a more than twofold change were considered as differentially expressed. All our microarray data sets are accessible through the GEO Series accession number GSE39869.

2.5.3 Microarray validation by qPCR

Ten genes (*CYP392A16* (*tetur06g04520*), *CYP392D2* (*tetur03g04990*), *CYP392D8* (*tetur03g05070*), *CYP392D10* (*tetur03g05110*), *tetur02g09840* (glycosyltransferase), *tetur16g03200* (MFS transporter), *tetur13g04550*, *tetur01g00490* (ID-RCD), *TuGSTd14* (*tetur29g00220*) and *tetur06g04970* (short chain reductase)) that were differentially expressed between both resistant strains (MR-VP, MAR-AB) and the susceptible strain were selected to confirm the microarray results by quantitative realtime PCR (qPCR). To further validate the biological importance of genes identified by microarray experiments, the expression levels of these 10 differentially expressed genes were also determined for a second independent susceptible strain (LS-VL). Mite selection, culture conditions, RNA isolation and DNAse treatment was performed as previously described for microarray experiments. First strand cDNA was synthesized from 2 µg of total RNA using Maxima first strand cDNA synthesis kit (Fermentas). Real-time PCR was done on a Mx3000p real-time PCR system (Stratagene) using Maxima SYBR green qPCR Master Mix (Fermentas) with 2-3 biological replicates and 3 technical replicates for each gene. Gene specific primers for the 10 differentially expressed genes and 2 housekeeping genes (actin, ribosomal protein 49) were designed using Primer 3 software (Rozen and Skaletsky 2000) (Table IV.1, primer names with "q" suffix). Relative levels of expression were calculated according to Pfaffl *et al*. (2001).

2.6 Clustering analyses

The GeneSpring GX11.0 software (Agilent Technologies) was used to perform a hierarchical clustering analysis of microarray expression data using the Pearson centered distance metric

and complete linkage rule. For gene clustering, the OrthoMCL (Van Dongen 2000) software version 2.0 with parameters (-v all -te 2 -scheme 7 -I 1.7) was applied with the species combination *Drosophila melanogaster, Tribolium castaneum, Acyrthosiphon pisum, T. urticae, Caenorhabditis elegans* and *Homo sapiens* In this analysis, a total of 103,935 sequences clustered into 13,876 gene families (76,810 genes in clusters; 27,125 singletons). Out of these, 3,069 clusters contained sequences from all six genomes. Of the protein-coding genes predicted for *T. urticae*, 11,831 were clustered in a total of 6098 groups. OrthoMCL uses a Markov clustering algorithm on a pre-calculated sequence similarity matrix to group (putative) orthologs and paralogs (Li *et al.* 2003). The matrix was built based on an all-against-all BLASTp ((Altschul *et al.* 1990), 2.2.24+; default parameters) and filtered according to the OrthoMCL manual. Where predicted, splice variants were removed from the data set (the longest protein sequence prediction was withheld). OrthoMCL clustering results of *T. urticae* proteins can be accessed at the *T. urticae* genome portal website (http://bioinformatics.psb.ugent.be/orcae/overview/Tetur).

A modified Fisher Exact p-value (EASE Score) was calculated (Huang *et al.* 2009) to measure the gene-enrichment in OrthoMCL clusters of our microarray expression data.

2.7 Signal peptide prediction

The presence of a signal peptide in all protein sequences investigated in this study was predicted with SignalP 3.0 (Bendtsen *et al.* 2004) using Hidden Markov Models (HMM) and Neural Networks (NN). Protein sequences were considered to have a signal peptide under the condition that both models predicted a signal peptide.

2.8 Gene family analysis

2.8.1 Intradiol ring-cleavage dioxygenases (ID-RCDs)

Gene-specific *T. urticae* ID-RCD primers (Table IV.1, primers with "s" suffix) were designed (using Primer 3 Software (Rozen and Skaletsky 2000)) in order to amplify a 600-850 bp fragment of ID-RCDs of the closely related spider mite *T. evansi*. *T. evansi* genomic DNA was extracted using a phenol chloroform extraction method as described by Khajehali et al. (2011). PCRs were performed in 50 µl reaction volumes with 38.7 µl distilled water; 5 µl 10x PCR buffer, 2 mM $MgCl_2$, 0.2 mM dNTP, 0.2 µM of each primer; 2 µl of template DNA (± 100 ng/ul) and 0.3 µl of Taq polymerase (Invitrogen) and run on a Biometra Thermocycler Professional (Westburg). PCR conditions were as follows: 2' 94°C, 35 × (20" 94°C, 55" 49°C, 2' 72°C) and 5' 72°C. All PCR products were separated by electrophoresis

on a 1% agarose gel and visualized by ethidium bromide staining. Subsequently, PCR products were purified with the Cycle Pure Kit (Omega Bio-Tek) and cloned into the pJET1.2/blunt vector (Fermentas). After heat-shock transformation of *E. coli* (DH5α) cells, plasmid DNA was obtained by miniprep (using the Plasmid Mini Kit (Omega Bio-Tek)) and inserted fragments were sequenced with pJET1.2F and pJET1.2R primers by LGC Genomics. Sequences were deposited in the GenBank database (GenBank accession number JQ736355-JQ736359)**.**

The presence of ID-RCDs in *Panonychus citri, Metaseiulus occidentalis, Varroa destructor* and *Ixodes scapularis*) was determined through a tBLASTn search (Altschul *et al.* 1990) of *T. urticae* ID-RCD proteins sequences against the published transcriptome dataset of *P. citri* (EMBL-EBI accession number: ERP000885), the genome of *M. occidentalis* (Mocc_1.0 assembly), the genome of *V. destructor* (BRL_Vdes_1.0 assembly) and against the genome of *I. scapularis* (ASM20861v1 assembly).

All protein sequences of *T. urticae* ID-RCDs (Grbić *et al.* 2011) were used as queries in BLASTp (cut off E-value $\leq e^{-10}$) searches of the NCBI non-redundant protein database. This resulted in 280 unique hits with a protein length between 176 and 500 amino acids. In the dataset obtained, 4 genera were overrepresented, namely *Aspergillus* (48 sequences), *Streptomyces* (36 sequences), *Rhodococcus* (12 sequences) and *Rhizobium* (9 sequences). All redundant amino acid sequences from these genera were removed for further analysis. Of the resulting 213 protein sequences, 191 were recognized by Conserved Domain Database (Marchler-Bauer *et al.* 2011) as members of the intradiol dioxygenase like subgroup (cd03457) and 22 sequences as members of the intradiol dioxygenase superfamily (cl01383). Next, nineteen functionally characterized ("classical'") intradiol dioxygenases (cd03459-cd03464), belonging to bacteria and fungi, were added to the dataset. Finally, 17 *T. urticae* ID-RCDs and five *T. evansi* homologues (see above) were incorporated into the analysis (see Appendix IV-A for GenBank accession numbers). The amino acid sequence alignment was constructed using MUSCLE (Edgar 2004). Model selection was done with ProtTest 2.4 (Abascal *et al.* 2005) and according to the Akaike information criterion the model WAG+I+G+F was optimum for phylogenetic analysis. Finally, a maximum likelihood analysis was performed using Treefinder (Jobb *et al.* 2004), bootstrapping with 1000 pseudoreplicates (LR-ELW). Phylogenetic trees were visualized and edited using MEGA5 (Tamura *et al.* 2011) and CorelDraw X3 (Corel Inc.), respectively.

Table IV.1 - Primers used for *T. evansi* ID-RCD sequencing and qPCR confirmation of microarray results

Tetur ID	Gene name/family	Primer	Sequence (5'-3')	(q)PCR-product (nt)	T_m (°C)
tetur18g03590	Ribosomal protein 49	18g03590_q_F	CTTCAAGCGGCATCAGAGC	105	62.19
		18g03590_q_R	CGCATCTGACCCTTGAACTTC		62.09
tetur03g09480	Actin	03g09480_q_F	GCCATCCTTCGTTTGGATTTGGCT	113	69.94
		03g09480_q_R	TCTCGGACAATTTCTCGCTCAGCA		69.43
tetur03g04990	CYP392D2	03g04990_q_F	TTTAAAATCAGCACAGGGTAAATTG	152	60.58
		03g04990_q_R	CACTAACTTGCTTCATGTTGTGTCTT		60.92
tetur03g05070	CYP392D8	03g05070_q_F	TGAGCTCAGAAACGCGAAT	100	59.69
		03g05070_q_R	CTCGATTTCATGGGTTGCTT		60.07
tetur03g05110	CYP392D10	03g05110_q_F	ATTGGATTCGAACGTCAACC	109	59.80
		03g05110_q_R	GTAAATTAAGAGGAGTGATTGTTGCT		58.48
tetur06g04520	CYP392A16	06g04520_q_F	AAATACCGAGGTCGGACGTA	117	59.45
		06g04520_q_R	AAGCACTTTTTCAATCTGGTCAC		59.69
tetur06g04970	Short chain reductases	06g04970_q_F	TGCTGGTCTTCATGGTTTCC	146	61.05
		06g04970_q_R	TGACTATGTTTCCTGCAGATTTGT		60.06
tetur29g00220	TuGSTd14	29g00220_q_F	CTTTGGCAGATCTCACCGTAA	119	60.26
		29g00220_q_R	GTCAGCGTAGTTAGTTTCCAGTTG		59.43
tetur02g09840	Glycosyltransferases	02g09840_q_F	GTTTACTCAGCAAATCCTCTTGC	104	59.45
		02g09840_q_R	TCTCGGTAAATTTTCCAATAATCTC		59.02
tetur16g03220	MFS	16g03220_q_F	ATATCCGTGATCAGTGCAACA	140	59.02
		16g03220_q_R	AATTGGAACTATTGGCACAGC		59.10
tetur01g00490	ID-RCD	01g04900_q_F	CCGAAAAGCTCACCAACATTCAAG	81	66.04
		01g04900_q_R	CGTTTCAAGTCATCGGGAAGAAAG		65.20
tetur13g04550	ID-RCD	13g04550_q_F	CTGGCAAGCCAATGCTTTA	100	59.96
		13g04550_q_R	ACCTCTGAGGAATCTTTCACCA		60.11
tetur01g00490	ID-RCD	01g00490_s_F	CCTTTGTCTTGTTTTCATTACCG	712	59.92
		01g00490_s_R	TGGATCAATGGCGACTGTG		61.71
tetur04g08620	ID-RCD	04g08620_s_F	TCCGATCCCGATTATGTCTC	686	59.85
		04g08620_s_R	CTGGTAGGCTCAATACCAAGTG		58.79
tetur06g00460	ID-RCD	06g00460_s_F	GAAAGACCTGGTAAATTCGTTG	680	57.80
		06g00460_s_R	TCAGGATCAATTCCCAAAGTG		59.92
tetur07g02040	ID-RCD	07g02040_s_F	GCCTGTTGATTACTTCTTCCTTG	737	59.32
		07g02040_s_R	GGCCCACATTAAATTTGACCTATG		62.63
tetur07g05940	ID-RCD	07g05940_s_F	CCATTCCGAAAGATCCATTG	660	60.27
		07g05940_s_R	GCACTGTAACCATTACCTTCTGG		59.94
tetur12g04671	ID-RCD	12g04671_s_F	GATCGTACAATAGTTGATTGTGCTC	624	59.14
		12g04671_s_R	TCTGAAACCATTTTCCTGTGG		59.96
tetur19g03360	ID-RCD	19g03360_s_F	CGTTATTGTTACAGCCGATCC	831	59.49
		19g03360_s_R	GGATCGATACCGAATATCATGG		60.38
tetur20g01160	ID-RCD	20g01160_s_F	CGTACTGATTAGCCACCAATCC	730	60.74
		20g01160_s_R	AGATGTCATTCCCTCGGTTG		59.93
tetur20g01790	ID-RCD	20g01790_s_F	TTGGCTGTTACTTTTGGTGAC	812	57.80
		20g01790_s_R	GAACCAAAGTATCCTGTTTCTCG		59.19
tetur28g01250	ID-RCD	28g01250_s_F	TTTCCACTCATTCAGAGAGACC	710	58.38
		28g01250_s_R	TGTTAAAGTAGTGACTGTTGGATCG		59.65
tetur44g00140	ID-RCD	44g00140_s_F	TGAACCCAGGCCTATTGATG	850	60.85
		44g00140_s_R	TCCGAGTGTAATGATTCCTTG		57.69

2.8.2 Lipocalins

Pfam domain searches of our microarray gene expression data revealed that several significantly up- and downregulated genes contained the lipocalin signature (PF08212.7, PF00061.18). Protein sequences of these genes were used as query in BLASTp searches (Altschul *et al.* 1990) of the NCBI non-redundant protein database. Each BLASTp search resulted mainly in hits with the highest bitscore for apolipoprotein D proteins of mammals. Next, a reference apolipoprotein D protein sequence of *Homo sapiens* (GenBank accession number: P05090) was used as query in BLASTp (cut off E-value $\leq e^{-5}$) against the proteome of *T. urticae* at the ORCAE website (http://bioinformatics.psb.ugent.be/orcae/overview/Tetur). Finally, the resulting hits were used as query in BLASTp (cut off E-value $\leq e^{-5}$) against the proteome of *T. urticae*. Using

this approach we identified 67 lipocalin candidates. Pseudogenes and gene fragments were separated from putative full-length lipocalins (58, grouped via OrthoMCL into clusters 10134, 10107, 19721, 19288 and 21421). The latter were together with a selected reference dataset of lipocalin genes ((Ganfornina *et al.* 2006), see Appendix IV-B for GenBank accession numbers) aligned using the profile alignment mode of Clustal X and the alignment of Sanchez et al. (Sanchez *et al.* 2003) as profile (Anderson *et al.* 2011). Tick lipocalin protein sequences were not included in our phylogenetic analysis in order to decrease the risk of long-branching artefacts (Ganfornina *et al.* 2006). Model selection was performed with ProtTest 2.4 (Abascal *et al.* 2005) and according to the Akaike information criterion the model WAG+G+F was optimum for phylogenetic analysis.

Finally, a maximum likelihood analysis was performed using Treefinder (Jobb *et al.* 2004), bootstrapping with 500 pseudoreplicates (LR-ELW). Phylogenetic trees were visualized and edited using MEGA5 (Tamura *et al.* 2011) and CorelDraw X3 (Corel Inc.), respectively. Secondary structures of *T. urticae* lipocalins were predicted using Jpred 3 (Cole *et al.* 2008) while GPI-anchor sites were predicted using PredGPI (Pierleoni *et al.* 2008).

2.8.3 Major facilitator superfamily (MFS) transporters

Pfam domain searches of our microarray expression data revealed the up- and downregulation of genes containing the MFS signature (PF07690.11). Most differentially expressed MFS genes grouped into three OrthoMCL clusters: 10032, 10082 and 10236. To determine the MFS class of genes in these clusters, protein sequences were used as query in BLASTp in the Transporter Classification DataBase (Saier *et al.* 2009). Transmembrane regions were predicted using TMHMM Server v. 2.0. (http://www.cbs.dtu.dk/services/TMHMM/).

3 Results

3.1 Host plant shift effects on gene expression

To examine genome-wide patterns of gene expression in *T. urticae*, we constructed an expression microarray (using the Agilent eArray platform, see Materials and Methods) with long oligonucleotide probes against all predicted genes of the London reference strain. We then used this array to examine expression changes associated with host plant change, as well as expression patterns in acaricide resistant strains. For the host plant change experiment, we transferred 1-3 day old females (London strain) from their common host, *Phaseolus vulgaris* (kidney beans), to a more challenging and less accepted host, *Solanum lycopersicum* (tomato). We used young females because this stage actively disperses with the wind to escape kin competition and over-exploitation (Kennedy and Smitley 1985; Bitume *et al.* 2011), and hence are expected to encounter potentially less favorable plants, on which they must immediately feed to produce eggs for colony establishment (colonies can then persist for many generations). We followed transcriptional changes over the short term to understand the initial responses, as well as after five generations on the new host. Briefly, female mites grown on beans were transferred to tomato, and transcriptional responses of mites were assessed at two hours (Tomato-2h), twelve hours (Tomato-12h) and after propagation for five consecutive generations (Tomato-5G). As assessed by the number of differentially expressed genes ($|\log_2(\text{Fold Change (FC)})| \geq 1$, Benjamini-Hochberg false discovery rate (FDR) < 0.05), the transcriptional response increased with time. Thirteen and 416 genes were differentially expressed after 2 and 12 hours, respectively, while 1,206 or about 7.5% of all predicted genes with probes on the array were differentially expressed after 5 generations (Fig. IV.1*A*). There was little overlap between genes associated with the early responses (Tomato-2h and Tomato-12h) and those with changed expression after 5 generations (Tomato-5G) (i.e., only 8.3 % of Tomato-5G was shared with Tomato-2h and Tomato-12h) (Fig IV.1*A*).

3.2 Acaricide-resistance effects on gene expression

To relate patterns of response between host adaptation and evolved pesticide resistance, we assessed gene expression patterns between two highly resistant field collected strains (MR-VP, MAR-AB) and the reference susceptible London strain (Khajehali *et al.* 2011). These field strains, one collected on beans and the second collected on roses (see Materials and Methods), are resistant when grown on bean, the host we used for assessing transcriptome variation among the three strains. As compared to the London strain by array hybridization,

we observed differences in transcript levels ($|\log_2(FC)| \geq 1$, FDR < 0.05) for 893 and 977 genes for MR-VP and MAR-AB, respectively (Fig. IV.1*B*). Our earlier work has shown that mite strains can be genetically diverse (Grbić *et al.* 2011; Van Leeuwen *et al.* 2012), potentially confounding comparison of gene expression across strains (polymorphisms can affect array-hybridization). However, our long oligo array is expected to be relatively robust to SNP and small indel changes (Hughes *et al.* 2001); more importantly, we validated with quantitative PCR (qPCR) (Fig. IV.2) a subset of genes predicted from the array to be differentially expressed between the London strain and a second susceptible strain (LS-VL, Van Leeuwen *et al.* 2005) with a different genetic background.

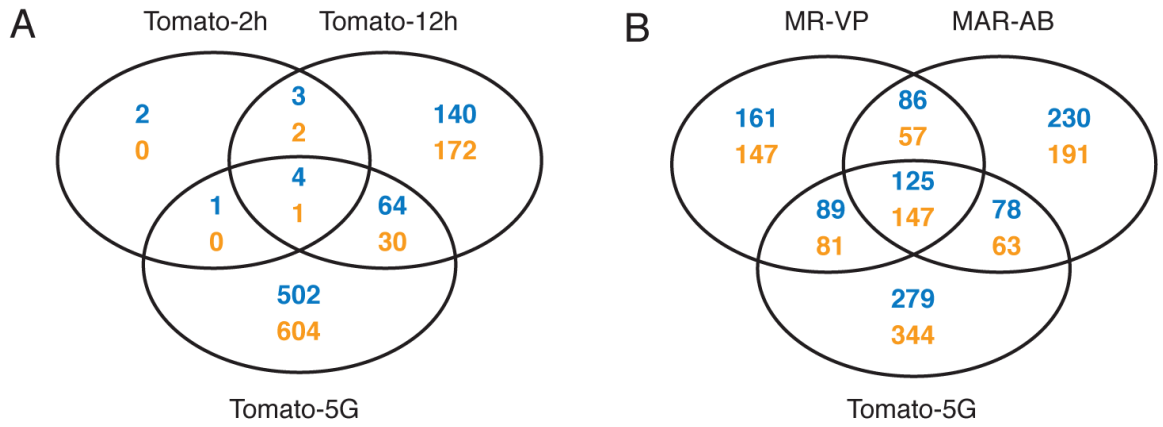


Figure IV.1 - Venn-diagrams depicting overlap among differentially expressed genes ($|\log_2(FC)| \geq 1$, FDR < 0.05) from pairwise comparisons of mites shifted from bean to tomato and of resistant mites. Blue: upregulated genes; orange: downregulated genes. (A) Comparisons for shift to tomato for 2h, 12h and five generations and (B) strains MAR-AB, MR-VP and Tomato-5G.

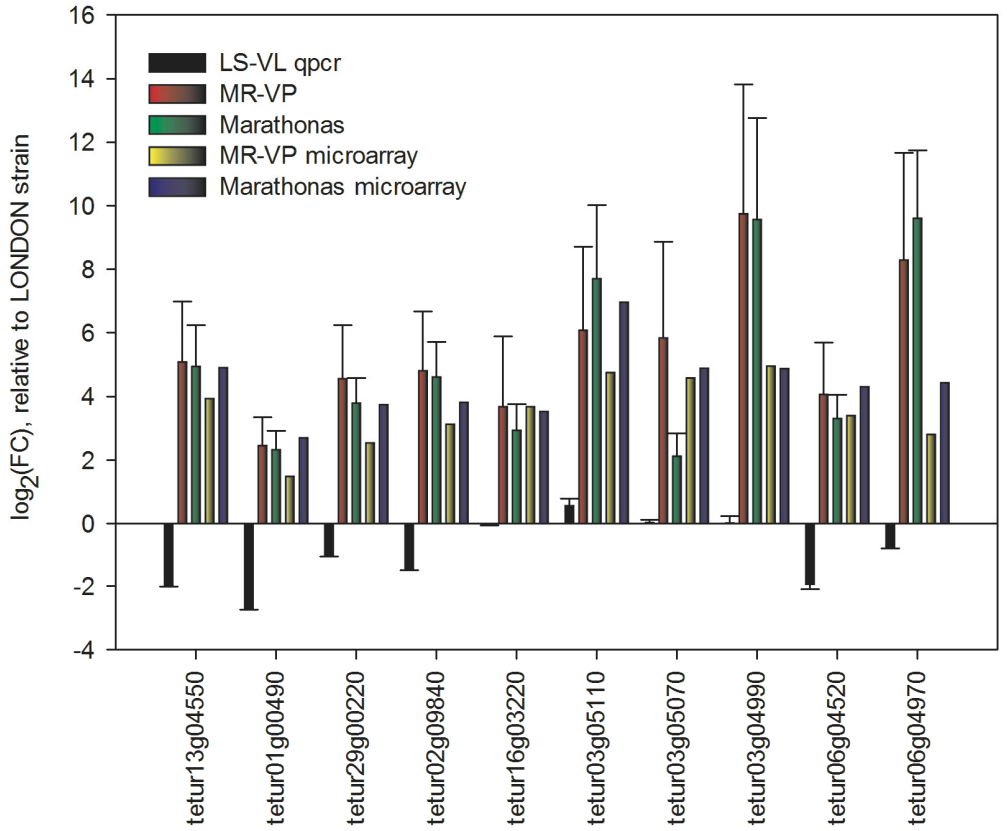


Figure IV.2 - Microarray validation by quantitative real-time PCR.

Validation was performed for 10 differentially expressed genes (*CYP392A16* (*tetur06g04520*), *CYP392D2* (*tetur03g04990*), *CYP392D8* (*tetur03g05070*), *CYP392D10* (*tetur03g05110*), *tetur02g09840* (glycosyltransferase), *tetur16g03200* (MFS transporter), *tetur13g04550*, *tetur01g00490* (ID-RCDs), *TuGSTd14* (*tetur29g00220*) and *tetur06g04970* (short chain reductase) for two susceptible strains (London and LS-VL) and both resistant strains (MR-VP, MAR-AB). Error bars represent the standard error of the calculated mean based on three biological replicates. Microarray expression data (MR-VP and MAR-AB, microarray) from this selection of genes is shown next to their qPCR expression data.

3.3 Relationships among transcriptome profiles

Although the resistant strains are genetically unrelated, there was an overlap of 415 differentially expressed genes (46.5% and 42.5 % of the total number of differentially expressed genes in MR-VP and MAR-AB, respectively). Further, we found that 49.5 % and 42.3 % of differentially expressed genes in the resistant MR-VP and MAR-AB strains were also differentially expressed after the London strain was transferred to tomato for five generations. A scatter plot of the fold changes for the intersection of differentially expressed genes between the host transfer and resistance datasets revealed a high correlation for gene expression levels (Spearman correlation: rho= 0.740, $p < 0.001$), Fig. IV.3*A*).

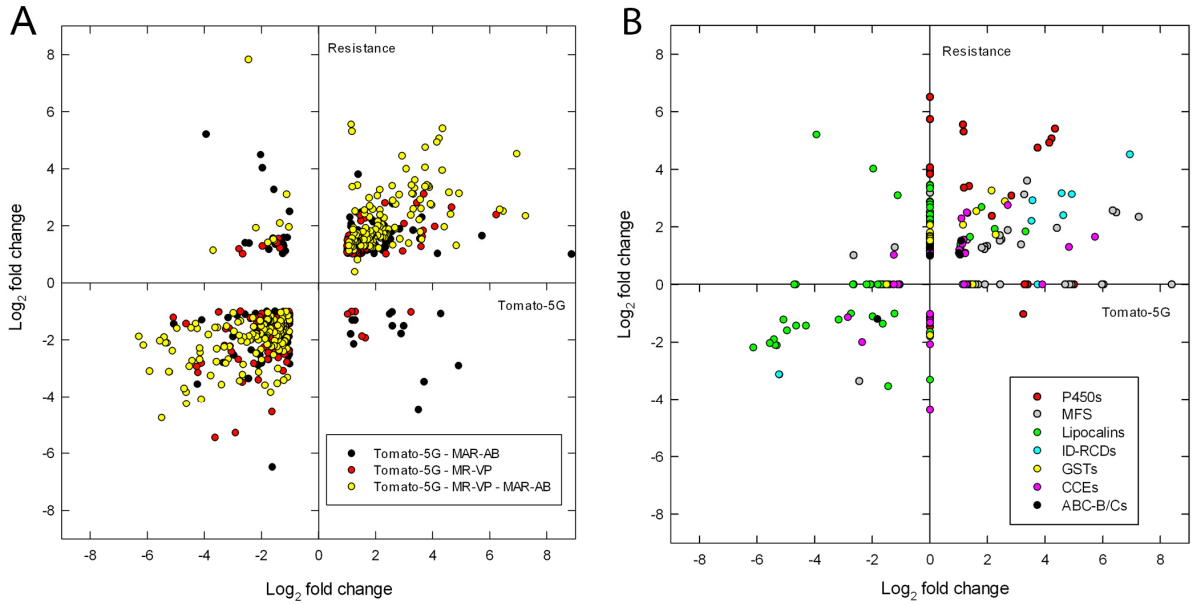


Figure IV.3 - Global changes in gene expression of two multi-resistant *T. urticae* strains (MR-VP and MAR-AB) relative to the London susceptible strain, compared to gene expression changes upon host plant change (Tomato-5G).

(*A*) Commonly differentially expressed genes ($|\log_2(FC)| \geq 1$, FDR < 0.05) in 2 multi-resistant strains (MR-VP and/or MAR-AB: "Resistance") and after host plant change for 5 generations (Tomato-5G): black=differentially expressed genes in Tomato-5G and MAR-AB, red= differentially expressed genes in Tomato-5G and MR-VP, yellow=differentially expressed genes in Tomato-5G, MR-VP and MAR-AB (the $\log_2$ of the average of fold changes of commonly differentially expressed genes of MR-VP and MAR-AB is plotted). (*B*) Fold changes of differentially expressed genes ($|\log_2(FC)| \geq 1$, FDR < 0.05), known to be implicated in detoxification and transport, in 2 multi-resistant strains (MR-VP and/or MAR-AB: "Resistance") and after host plant change for 5 generations (Tomato-5G): red=P450 monooxygenases (P450s), black=ATP-binding cassette transporters, class B and C (ABC-B/Cs), green=lipocalins, pink=carboxyl-cholinesterases (CCEs), yellow=glutathione S-transferases (GSTs), light blue=intradiol ring-cleavage dioxygenases (ID-RCDs), gray=MFS transporters (OrthoMCL clusters 10032, 10082 and 10236)

Further, hierarchical clustering (Pearson centered distance metric, complete linkage rule) across all the expression data revealed that expression patterns for the two resistant strains and for mites feeding on tomato for five generations clustered together, and not with early responses for mites transferred to tomato for 2h or 12 h. The overlap between resistance and host-plant change was even more striking when genes were grouped in gene (sub)families by OrthoMCL clustering (Table IV.2). Shared responses were largely mediated by a few gene families, and in some cases a large proportion of all family members were included. As revealed by PFAM-domains searches of OrthoMCL clusters (48% of genes in shared OrthoMCL clusters (Table IV.2) have an assigned PFAM-domain with E-value $\leq e^{-5}$), some responsive families belong to those that have been commonly implicated in detoxification or transport of xenobiotics (e.g. CCEs, P450s, GSTs and ABC-transporters (ABC-B/Cs)). Among these, P450 genes stood out as being markedly differentially expressed among resistant mites and after host transfer of the susceptible London strain (Fig. IV.3*B*).

Intriguingly, some of the most strongly affected gene families in both experiments have signatures that have, until now, not been commonly associated with response to xenobiotics in arthropods. To shed insights into the *T. urticae* polyphagous life history, we therefore examined the composition and the nature of transcriptional responses for such families of moderate size (ten members or more, see below). We note, however, that genes of unknown function had some of the most striking expression changes. For many such genes, encoded products are predicted to be secreted (Table IV.2). An example is OrthoMCL cluster 10257 for which *tetur11g05420*, *tetur11g05450* and *tetur46g00020* were upregulated by ~700-fold upon host transfer to tomato for 5 generations.

3.4 Intradiol ring-cleavage dioxygenases

A set of 17 genes encoding secreted proteins identified as intradiol ring-cleavage dioxygenases (ID-RCDs) were among the most striking differentially expressed in our analysis (Table IV.2). These genes, belonging to the 'intradiol dioxygenase like' subgroup (cd03457) according to the Conserved Domain Database (Marchler-Bauer *et al.* 2011), were recently identified as a case of lateral gene transfer in the genome of *T. urticae*, and have not been reported in other metazoan genomes to date (Grbić *et al.* 2011). More than half of the genes in this family were differentially expressed upon host plant change and in multi-resistant strains, and their expression patterns were highly correlated (Fig. IV.3*B*, Fig. IV.4).

Table IV.2 - The extent of gene expression changes within gene clusters.

Gene clusters are as defined by OrthoMCL clustering. Percentage of genes differentially expressed within each OrthoMCL cluster in MR-VP, MAR-AB and Tomato-5G (Tom-5G) are shown. Only shared OrthoMCL clusters (≥ 10 genes) where at least 20% of members are differentially expressed in either MR-VP, MAR-AB or Tomato-5G are shown. Gene family names have been linked to OrthoMCL clusters when possible. Many clusters consist of proteins with no PFAM-hits (hypothetical proteins). OrthoMCL clusters are sorted based on the average of the percentage of differentially expressed genes within each OrthoMCL cluster across MR-VP, MAR-AB and Tomato-5G. For each OrthoMCL cluster the percentage of genes predicted with a signal peptide was calculated using SignalP 3 (Bendtsen *et al*. 2004). Clusters mentioned in this study are shaded in grey.

OrthoMCL	Members	% S_p[1]	Gene Family	PFAM domain(s)	MR-VP[2]	MAR-AB[2]	Tom-5G[2]
10134	25	80.0	lipocalins	PF08212.7, PF00061.18	68.0*	68.0*	64.0*
10364	16	43.8	low-density lipoprotein receptors	PF00057.13	62.5*	56.3*	68.8*
10287	17	100.0	intradiol ring-cleavage dioxygenases	PF00775.16	52.9*	52.9*	70.6*
10012	52	5.8	P450 monooxygenases	PF00067.17	51.9*	65.4*	36.5*
11149	10	90.0	hypothetical proteins	-	50.0*	30.0	70.0*
11121	10	40.0	hypothetical proteins	-	10.0	70.0*	60.0*
10289	18	83.3	hypothetical proteins	-	27.8*	44.4*	66.7*
10107	30	100.0	lipocalins	PF08212.7, PF00061.18	43.3*	26.7*	66.7*
10176	24	45.8	hypothetical proteins	-	50.0*	50.0*	29.2*
10422	15	53.3	hypothetical proteins	-	33.3*	40.0*	40.0*
10254	19	0.0	Kelch-related proteins	PF00651.26, PF07707.10	52.6*	31.6*	26.3*
10475	14	92.9	hypothetical proteins	-	21.4	64.3*	21.4
10257	19	78.9	hypothetical proteins	-	31.6*	26.3*	47.4*
10236	20	5.0	MFS	PF07690.11	15.0	40.0*	50.0*
10115	16	0	δ GSTs	PF00043.20,PF13417.1	12.5	56.3*	31.3*
10041	29	3.4	short-chain reductases	PF13561.1, PF00106.20	13.8	31.0*	51.7*
10032	53	7.5	MFS	PF07690.11	18.9*	32.1*	45.3*
10420	15	93.3	hypothetical proteins	-	20.0*	46.7*	26.7
10831	11	81.8	cysteine proteases	PF08127.8, PF00112.18	18.2	27.3	45.5*
10689	12	41.7	cuticular proteins	PF00379.18	16.7	41.7*	25.0
10040	51	0	NYN-domain family	PF01936.13	29.4*	21.6*	31.4*
10008	11	0	P450 monooxygenases	PF00067.17	9.1	36.4*	36.4*
10841	11	0	hypothetical proteins	DUF1768	27.3	36.4*	18.2
10085	27	74.1	alpha/beta hydrolase fold-family	PF04083.11, PF00561.15	29.6*	18.5	33.3*
10204	21	95.2	hypothetical proteins	-	14.3	19.0	47.6*
10157	25	0	hypothetical proteins	-	40.0*	12.0	28.0*
10158	24	58.3	carboxyl/cholinesterases	PF00135.23	12.5	29.2*	37.5*
10074	40	67.5	carboxyl/cholinesterases	PF00135.23	22.5*	30.0*	25.0*
10235	20	90.0	cysteine proteases	PF08246.7, PF00112.18	15.0	10.0	50.0*
10491	14	35.7	hypothetical proteins	-	28.6*	21.4	21.4
10010	122	18.9	PAN-domain family	PF00024.21	15.6*	18.9*	36.1*
10149	10	20.0	short chain reductases	PF00106.20	10.0	30.0	30.0
10199	20	80.0	asparaginyl peptidases	PF01650.13	10.0	30.0*	25.0
10094	34	8.8	hypothetical proteins	-	8.8	38.2*	17.6
10225	11	0.0	μ GSTs	PF00043.20, PF02798.15	18.2	27.3	18.2*
10200	22	0.0	hypothetical proteins	DUF3421	13.6	27.3*	22.7
10066	42	54.8	hypothetical proteins	-	19.0*	23.8*	19.0*
10014	102	1.0	glycosyl transferases	PF00201.13, PF04101.11	10.8	21.6*	28.4*
10082	36	2.8	MFS	PF07690.11	16.7	22.2*	16.7
10003	239	1.3	F-box proteins	PF12937.2	19.2*	10.9*	21.3*

[1] percentage of gene-members predicted with a signal peptide using SignalP 3.0 (Bendtsen *et al*. 2004)
[2] OrthoMCL clusters with an EASE (modified Fisher exact p-value) score < 0.05 are indicated with an asterisk (Huang *et al*. 2009)

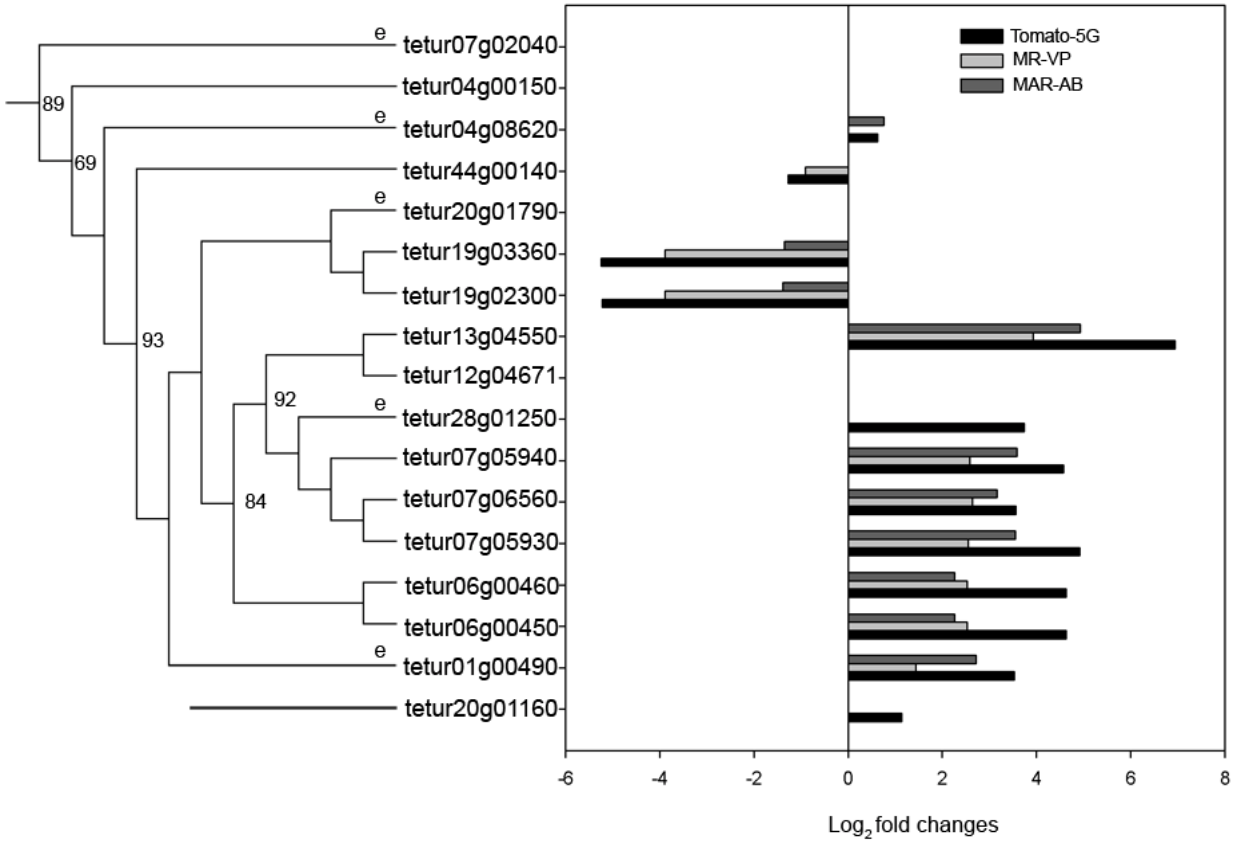


Figure IV.4 - Phylogenetic relationship of *T. urticae* ID-RCDs linked to expression levels (log_2(FC)) in acaricide multi-resistant strains (MR-VP and MAR-AB) and after host plant shift to tomato (S. *lycopersicum*) for five generations.

Genes with detected orthologues in *T. evansi* are depicted with an 'e'

ID-RCDs catalyze the oxygenolytic fission of catecholic substances, allowing bacteria and fungi to degrade aromatic rings, a crucial step in the global carbon cycle. Although bacteria usually harbor only one to four ID-RCD genes, this family has proliferated in *T. urticae* to 16 complete ID-RCDs and a pseudogene (*tetur07g06560*). Spider mite ID-RCDs share the conserved 2 His 2 Tyr non-heme iron (III) active site with previously described and functionally characterized ID-RCDs (such as catechol, hydroxyquinol and protocatechuate ID-RCDs) (Vetting and Ohlendorf 2000; Ferraroni *et al.* 2005; Vaillancourt *et al.* 2006; Matera *et al.* 2010; Grbić *et al.* 2011) (Fig. IV.5). They are distributed over eleven genomic scaffolds and all but one (*tetur07g02040*) are intronless. Clusters of duplicated *T. urticae* ID-RCD genes were found on several scaffolds (Fig. IV.4).

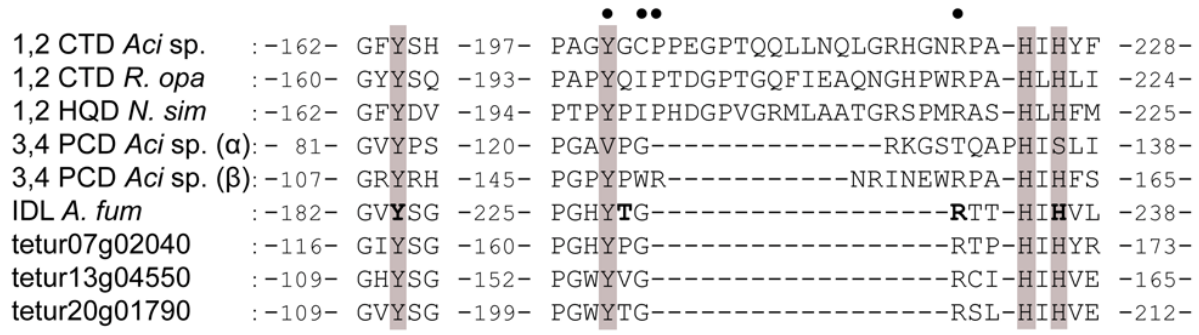


Figure IV.5 - Alignment of conserved residues in "classical" and secreted ID-RCDs.
CTD: catechol ID-RCD; PCD: protocatechuate ID-RCD; HQD: hydroxyquinol ID-RCD; IDL: intradiol dioxygenase like (cd03457); Aci: *Acinetobacter* sp.*;* R. opa: *Rhodococcus opacus*; N. sim: *Nocardia simplex*; and A. fum: *A. fumigatus*. Tetur07g02040, tetur13g04550 and tetur20g01790 are ID-RCD representatives of *T. urticae*. The 2 His 2 Tyr non heme iron (III) binding sites are indicated by shading. Residues defined by crystallographic (Vetting and Ohlendorf 2000; Ferraroni *et al.* 2005; Matera *et al.* 2010) studies to have an influence on substrate interaction in "classical" ID-RCDs (CTD, PCD and HQD) are indicated by black dots. The predicted binding residues of epicatechin in the protein sequence of *A. fumigatus* are indicated in bold (Roopesh *et al.* 2012).

We detected by PCR five orthologous genes in the closely related species, *T. evansi*, an oligophagous specialist of Solanaceae (Fig. IV.4 and IV.5). We also found ID-RCD sequences (E-value $\leq 2e^{-36}$) in the RNAseq data from the citrus red spider mite, *Panonychus citri* (EMBL-EBI accessionnumber: ERP000885). However, we found no trace of their presence in the genomes of other, non plant-feeding Acari, such as *M. occidentalis (*a predatory mite), *V. destructor* (an ectoparasite of bees), or *I. scapularis* (a blood-feeding tick). This suggests that a horizontal transfer occurred after the split of these lineages from the Tetranychidae. Phylogenetic analysis revealed that spider mite ID-RCDs cluster with a

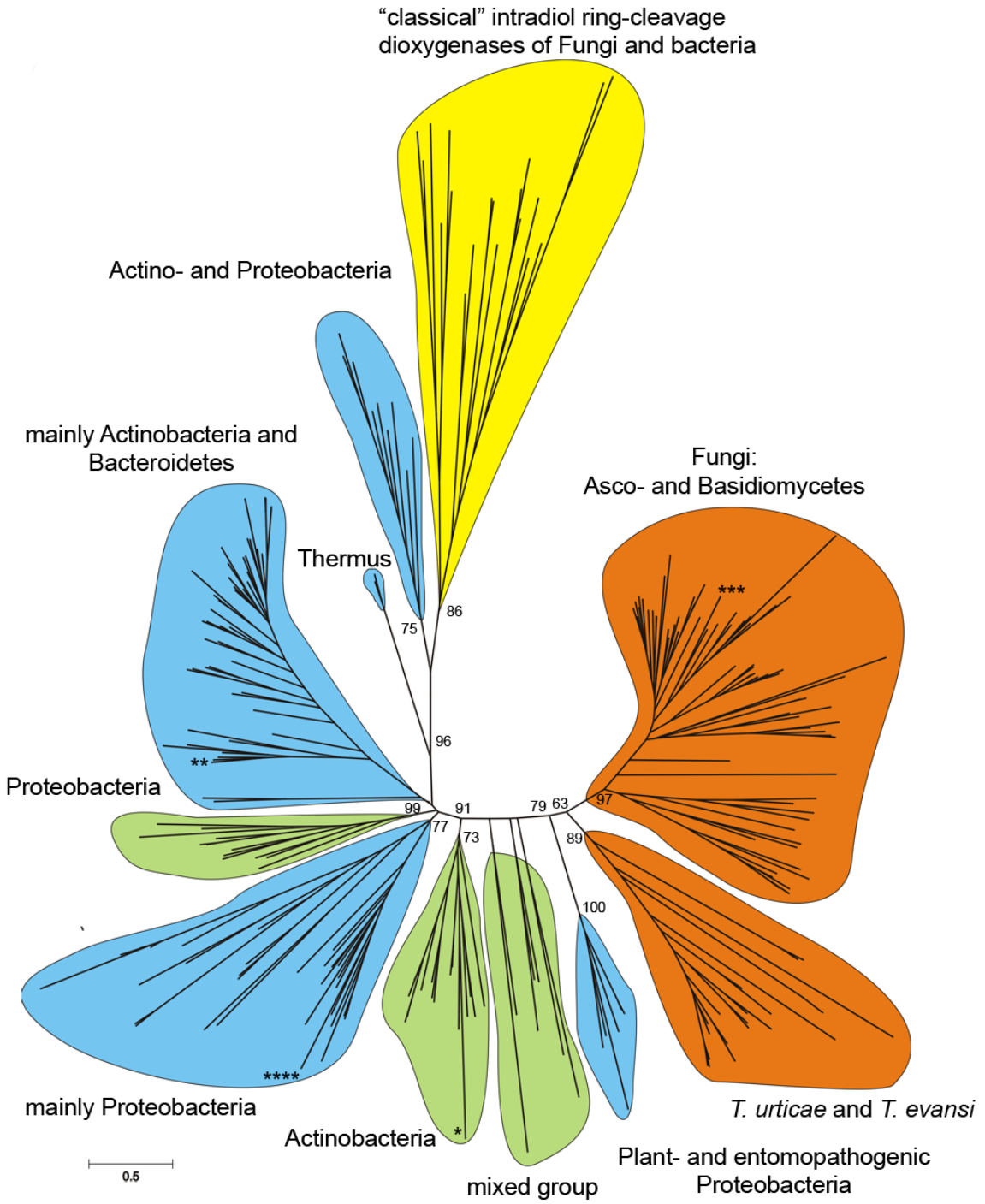


Figure IV.6 - Maximun likelihood unrooted tree of 17 ID-RCDs of *T. urticae* (and five *T. evansi* orthologues) with 232 bacterial and fungal sequences.

Color codes indicate the percentage of secretion within the clade: yellow =not secreted, blue < 55 %, green 55-85%, and orange > 85% secreted. All members of the *T. urticae* clade are secreted and form a sister clade to fungal secreted dioxygenases, sharing a most recent common ancestor with plant and entomopathogenic Proteobacteria. The "classical" biochemically characterized ID-RCDs (CTD, PCD and HQD) are not secreted and cluster together as an outgroup. The phylogenetic positions of the ID-RCD protein sequences of *Naegleria gruberi* (Protozoa), *Schistosoma mansoni* (Metazoa), *Phytophtora. infestans* (oomycete) and *Haloferax volcanii* (Archaea) are indicated by *, **, *** and ****, respectively.

group of fungal ID-RCDs that share a common ancestor with plant and entomopathogenic bacterial ID-RCDs (such as *Xenorhabdus* sp. and *Photorhabdus* sp. (Goodrich-Blair and Clarke 2007)) (Fig. IV.6). None of these fungal ID-RCDs belong to previously characterized "classical" fungal and bacterial ID-RCDs known to metabolize well-characterized substrates such as catechol and protocatechuate. In contrast to the characterized cytoplasmic enzymes, this large clade of predicted secreted forms of ID-RCDs (Fig. IV.6, Appendix IV-A) has not yet been recognized and thoroughly characterized, although we found these proteins in proteomic data on fungal secretomes (Paper *et al.* 2007; Yang *et al.* 2012).

3.5 Lipocalins

We also found that lipocalins, small proteins capable of binding to hydrophobic molecules, were strongly differentially expressed (68% of genes of OrthoMCL cluster 10134) between London and resistant strains, and dynamically over time to host plant change (Table IV.2, Fig IV.7*A*). As revealed by hierarchical clustering, lipocalin expression patterns of resistant strains and mites feeding on tomato for five generations group together. Some of these lipocalin genes are strongly and progressively induced in mites feeding for two and twelve hours on tomato, but are completely down regulated after five generations, while stable induced expression is maintained for other lipocalin genes (Fig. IV.7*A*). Of the 58 complete lipocalins we annotated in the *T. urticae* genome, half are concentrated on only three scaffolds (twenty on scaffold 6, eight on scaffold 1, and five on scaffold 31) (Table IV.3). The number of *T. urticae* lipocalins far exceeds those reported in insects (*D. melanogaster*: 4, *Apis mellifera*: 4, *Rhodnius prolixus*: 22) (Ganfornina *et al.* 2006) and in humans (10, (Breustedt *et al.* 2006)) but is in the same range as in ticks (Ribeiro *et al.* 2006). Thirty-six (62.0%) *T. urticae* lipocalins were predicted to have an anti-parallel β-barrel, while 15 (25.9%) had only a small deviation from the canonical lipocalin secondary structure (Flower *et al.* 2000) (Table IV.3). *T. urticae* lipocalins do not have a GPI-anchor signal omega site and are, with the exception of *tetur31g00780*, predicted to have a signal peptide (Table IV.3).

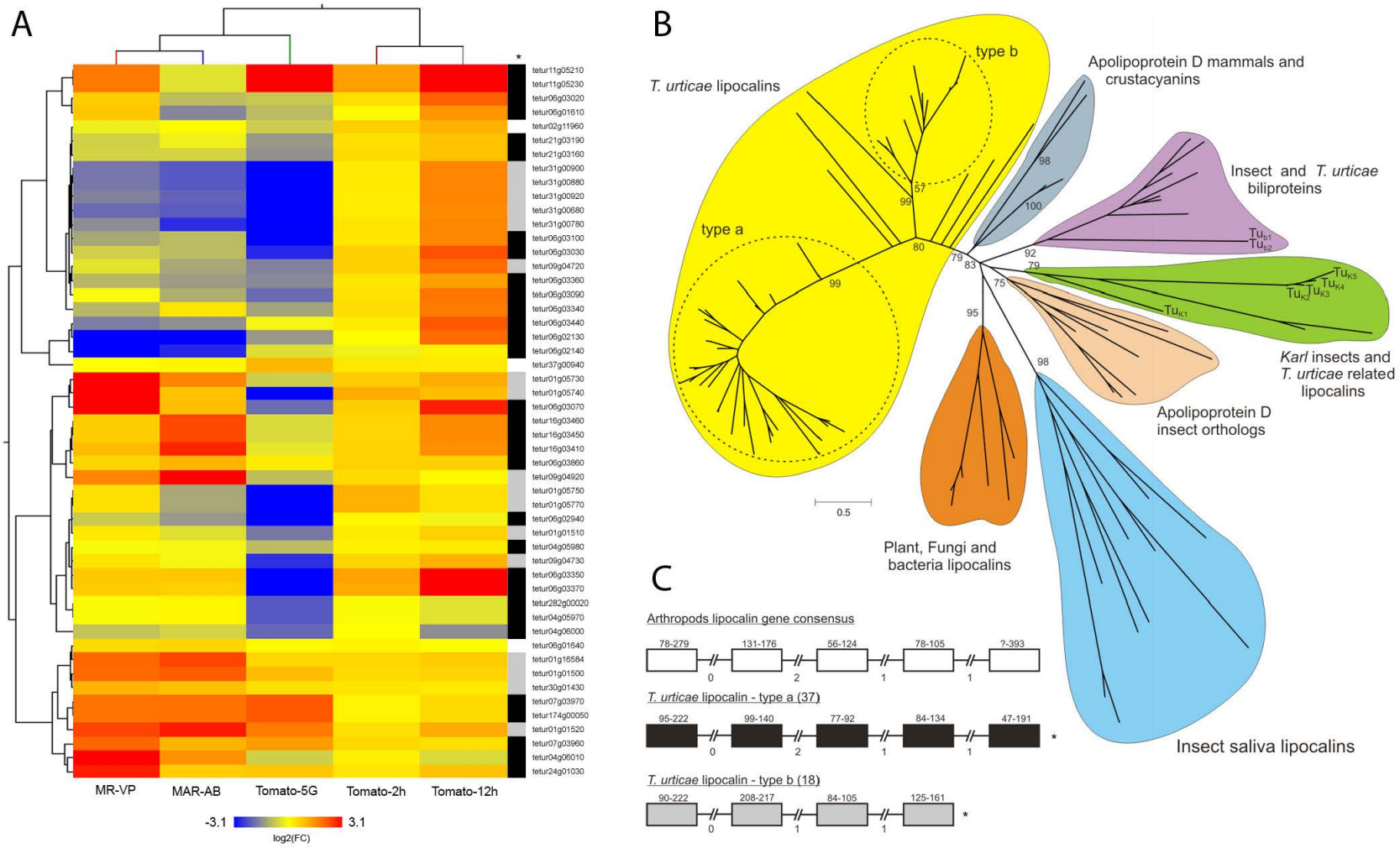

A
MR-VP
MAR-AB
Tomato-5G
Tomato-2h
Tomato-12h
-3.1
3.1
log2(FC)
B
T. urticae lipocalins
type b
type a
Apolipoprotein D mammals and crustacyanins
Insect and T. urticae biliproteins
Karl insects and T. urticae related lipocalins
Apolipoprotein D insect orthologs
Plant, Fungi and bacteria lipocalins
Insect saliva lipocalins
0.5
C
Arthropods lipocalin gene consensus
78-279
131-176
56-124
78-105
?-393
T. urticae lipocalin - type a (37)
95-222
99-140
77-92
84-134
47-191
T. urticae lipocalin - type b (18)
90-222
208-217
84-105
125-161

Figure IV.7

Figure IV.7 (continued) - Lipocalins in *T. urticae*

(*A*) Clustering of lipocalin gene expression between conditions (MAR-AB, MR-VP and host change (Tomato-2h, Tomato-12h and Tomato-5G)) reveals correlation of expression levels. Clustering was performed in GeneSpring GX11.0, using the hierarchical clustering algorithm with Pearson centered distance metric and complete linkage rule. The color bar with corresponding log_2(FC) values is show at the bottom of the figure. The bar under the asterisk indicates the number and phases of introns of each *T. urticae* lipocalin (as explained in panel C: black= type a, grey = type b, white=neither type a or b). (*B*) Maximun likelihood unrooted tree depicting the phylogenetic relationship of the expanded (58 genes) lipocalin family of *T. urticae*. Most of *T. urticae* lipocalins cluster with mammalian Apolipoprotein D and crustacyanins. *T. urticae* homologues of insect biliproteins and Karl of *D. melanogaster* are indicated as Tu_{b1} (tetur07g03790), Tu_{b2} (tetur174g00050) and Tu_{K1} (tetur02g11960), Tu_{K2} (tetur01g01510), Tu_{K3} (tetur01g1500), Tu_{K4} (tetur01g16584) and Tu_{K5} (tetur01g01520), respectively. Members within two sub-clades of the *T. urticae* lipocalins have similar numbers and phases of introns (type a or b, see panel C) and are depicted by circles. The highly divergent tick lipocalin protein sequences were not included in our phylogenetic analysis in order to decrease the risk of long-branch artefacts (Ganfornina *et al*. 2006) (*C*) Comparison of lipocalin gene structure consensus between arthropods (Sanchez *et al.* 2006) and *T. urticae*, revealing a new *T. urticae* gene structure (type b). Square boxes (not drawn to scale) represent exons. Numbers above boxes represent the exon size range (bp) while numbers between boxes represent intron phases.

Two main types of lipocalin gene organization were apparent. Thirty-seven genes have five exons and an intron phase pattern of 0-2-1-1, corresponding to the arthropod lipocalin gene consensus pattern, while 18 lipocalin genes have only 4 exons and a 0-1-1 intron phase pattern, a gene structure also reported for a moth lipocalin (Sanchez *et al.* 2006) (Table IV.3 and Fig. IV.7*C*). Most *T. urticae* lipocalins cluster together with a previously described clade (Ganfornina *et al.* 2000; Wade *et al.* 2009) comprising vertebrate apolipoprotein D (ApoD) and crustacyanins with a high bootstrap support (83%) (Fig. IV.7*B*, Table IV.3 and Appendix VI-B). Within this large *T. urticae* lipocalin clade, most lipocalins cluster according to their intron phase pattern (Fig. IV.7*B* and Fig IV.7*C*). However, five *T. urticae* lipocalins (tetur01g01500, tetur01g01510, tetur*01g01520*, tetur01g16584 and tetur02g09610) grouped together with Karl, a lipocalin expressed in the blood-cells of *D. melanogaster*, and two (tetur174g00050 and tetur07g03970) were closely related to insect biliproteins (Fig. IV.7*B*).

Table IV.3 - *T. urticae* lipocalin properties

Tetur ID[1]	β-strands/helices[2]	Intron phase pattern[3]	S_p [4]	Length (AA)	Strand	OrthoMCL
tetur01g01500	7+1/1+1	0_1_1	Y	190	+	19288
tetur01g01510	8+1/1+1	0_1_1	Y	195	+	19288
tetur01g01520	9+1/1+1	0_1_1	Y	187	+	19288
tetur01g05730	8+1/2+1	0_1_1	Y	208	+	10134
tetur01g05740	8+1/1+1	0_1_1	Y	192	+	10134
tetur01g05750	8+1/1+1	0_1_1	Y	187	-	10134
tetur01g05770	8+1/1+1	0_1_1	Y	187	+	10134
tetur01g16584	8+1/1+1	0_1_1	Y	195	+	10134
tetur06g03550	8/1+1	0_1_1	Y	192	+	10134
tetur09g04720	8+1/1+1	0_1_1	Y	198	+	10134
tetur09g04730	8+1/1+1	0_1_1	Y	195	+	10134
tetur09g04920	8+1/1+1	0_1_1	Y	194	-	10134
tetur30g01430	8+1/2+1	0_1_1	Y	218	-	10134
tetur31g00680	8+1/1+1	0_1_1	Y	199	-	10134
tetur31g00780	8+1/1+1	0_1_1	N	177	+	10134
tetur31g00880	8+1/1+1	0_1_1	Y	199	+	10134
tetur31g00900	8+1/1+1	0_1_1	Y	199	+	10134
tetur31g00920	8+1/1+1	0_1_1	Y	199	+	10134
tetur02g11960	8+2/1+1	0_2_1	Y	218	-	15855
tetur37g00940	8+1/1+1+1	0_2_1	Y	213	+	19721
tetur04g05970	8+1/1+1	0_2_1_1	Y	193	+	10107
tetur04g05980	8+1/1+1	0_2_1_1	Y	187	+	10107
tetur04g06000	8+1/1+1	0_2_1_1	Y	194	+	10107
tetur04g06010	8/1+1	0_2_1_1	Y	182	-	10107
tetur05g07070	8+1/1+1	0_2_1_1	Y	194	+	10134
tetur06g01610	8/1+1	0_2_1_1	Y	167	-	10107
tetur06g01640	8+1/1+1	2_1_1	Y	184	-	10134
tetur06g02130	8+1/1+1	0_2_1_1	Y	183	-	10107
tetur06g02140	8+1/1+1	0_2_1_1	Y	187	-	10107
tetur06g02670	8/1+1	0_2_1_1	Y	179	-	10107
tetur06g02940	2+1+1/2+2+1	0_2_1_1	Y	149	+	10107
tetur06g03020	8+2/1+1	0_2_1_1	Y	192	-	10107
tetur06g03030	3+1+1/1+2+1	0_2_1_1	Y	162	-	10107
tetur06g03070	3+1+1/2+2+1+1	0_2_1_1	Y	172	+	10107
tetur06g03090	8/1+1	0_2_1_1	Y	165	+	10107
tetur06g03100	8+1/1+1	0_2_1_1	Y	184	+	10107
tetur06g03340	8+1/1+1	0_2_1_1	Y	185	-	10107
tetur06g03350	8/1+1	0_2_1_1	Y	179	+	10107
tetur06g03360	8+1/1+1	0_2_1_1	Y	183	-	10107
tetur06g03370	8+1/1+1	0_2_1_1	Y	183	+	10107
tetur06g03440	8+1/1+1	0_2_1_1	Y	188	-	10107
tetur06g03530	8/2+2	0_2_1_1	Y	181	+	10107
tetur06g03860	8+1/1+1	0_2_1_1	Y	191	+	10134
tetur06g06691	8+1/1+1	0_2_1_1	Y	172	+	10107
tetur07g03960	9+1/1+1	0_2_1_1	Y	201	-	10134
tetur07g03970	8+1/1+1	0_2_1_1	Y	199	+	21421
tetur11g05210	8+1/1+1	0_2_1_1	Y	187	-	10107
tetur11g05230	8+1/1+1	0_2_1_1	Y	187	-	10107
tetur16g03410	8+1/1+1	0_2_1_1	Y	184	+	10107
tetur16g03450	8+1/1+1	0_2_1_1	Y	184	-	10107
tetur16g03460	8+1/1+1	0_2_1_1	Y	184	+	10107
tetur174g00050	8+1/1+1	0_2_1_1	Y	199	-	21421
tetur18g00900	8+1/1+1	0_2_1_1	Y	200	-	10107
tetur21g03160	7/1+1	0_2_1_1	Y	174	-	10107
tetur21g03190	8/1+1	0_2_1_1	Y	171	-	10107
tetur21g03340	8/1+1	0_2_1_1	Y	175	+	10107
tetur24g01030	8+1/3+1	0_2_1_1	Y	227	+	10134
tetur282g00020	8+1/1+1	0_2_1_1	Y	193	+	10107

[1] *T. urticae* accession numbers can be accessed through the ORCAE website (http://bioinformatics.psb.ugent.be/orcae/overview/Tetur)
[2] secondary structures were predicted using JPred3 (Cole *et al.* 2008), predicted b strands ("E") and helices ("H") were counted if they were at least three amino acids long and separated by at least 3 amino acids; "8+1/1+1" represents the typical repeated +1 topology β-barrel structure of lipocalins with a 3_{10}-like and α-helix at the N-and C-terminal end, respectively.
[3] the intron phase of each intron is separated by an underscore
[4] S_p: predicted with (Y) or without (N) a signal peptide using SignalP 3.0 (Bendtsen *et al.* 2004)

3.6 Major Facilitator Superfamily

Among the genes differentially expressed in both MR-VP, MAR-AB and Tomato-5G (Table IV.2), three OrthoMCL clusters (10082, 10032 and 10236) with a total of 109 genes shared the PFAM-domain PF07690.11 that characterizes the Major Facilitator Superfamily (MFS), also known as the uniporter-symporter-antiporter family. Members of OrthoMCL clusters 10236 and 10082 were highly similar (E-value $\leq e^{-17}$ and $\leq e^{-5}$, respectively) to the Anion/CationSymporter (ACS) family and the Na^+ dependent glucose transporter family, respectively (Table IV.4). On the other hand, most members of cluster 10032 showed similarity (E-value $\geq e^{-10}$, Table IV.4) to bacterial tetracycline:H+ antiporters and their mammalian homologs, the heme-carrier proteins/thymic-folate cotransporters (Shayeghi *et al.* 2005). All differentially expressed members of OrthoMCL cluster 10032 were upregulated in both MR-VP and MAR-AB, and 87.5 % of the differentially expressed members were also upregulated on tomato for five generations. *Tetur11g05410* was more than 300-fold upregulated by transfer to tomato. Moreover, 16 members of cluster 10032 were already upregulated in mites transferred from bean to tomato for 12 hours.

Table IV.4 - Classification of *T. urticae* MFS transporter genes (from OrthoMCL 10032, 10082 and 10236) determined by BLASTp in the Transporter Classification DataBase (Saier *et al.* 2009)

				10236			
Tetur ID[1]	length (AA)	TM[2]	S_p[3]	TCID name of best blastp hit at TCDB[4]	TCID	E-value	TCDB family
tetur04g02320	505	10	N	Sialin	2.A.1.14.10	e-52	(ACS) Family*
tetur05g04460	538	10	N	brain synaptic vesicle anion:Na^+ symporter	2.A.1.14.13	e-42	(ACS) Family
tetur08g06320	530	11	N	brain synaptic vesicle anion:Na^+ symporter	2.A.1.14.13	e-43	(ACS) Family
tetur08g06330	508	12	N	Sialin	2.A.1.14.10	e-52	(ACS) Family
tetur08g06340	527	11	N	Sialin	2.A.1.14.10	e-52	(ACS) Family
tetur08g06350	512	11	N	Sialin	2.A.1.14.10	e-46	(ACS) Family
tetur08g06360	503	11	N	Sialin	2.A.1.14.10	e-46	(ACS) Family
tetur08g06370	508	9	N	Sialin	2.A.1.14.10	e-55	(ACS) Family
tetur08g06410	512	11	N	Sialin	2.A.1.14.10	e-57	(ACS) Family
tetur08g06870	529	8	N	Sialin	2.A.1.14.10	e-58	(ACS) Family
tetur08g06890	527	9	N	Sialin	2.A.1.14.10	e-61	(ACS) Family
tetur09g02840	489	12	N	Sialin	2.A.1.14.10	e-43	(ACS) Family
tetur17g01270	522	11	N	Sialin	2.A.1.14.10	e-54	(ACS) Family
tetur17g01300	530	10	N	Sialin	2.A.1.14.10	e-50	(ACS) Family
tetur05g04570	353	5	N	Sialin	2.A.1.14.10	e-27	(ACS) Family
tetur05g04590	498	11	N	Sialin	2.A.1.14.10	e-36	(ACS) Family
tetur05g04600	280	6	N	brain synaptic vesicle anion:Na^+ symporter	2.A.1.14.13	e-17	(ACS) Family
tetur08g06400	447	7	N	brain synaptic vesicle anion:Na^+ symporter	2.A.1.14.13	e-45	(ACS) Family
tetur60g00010	240	6	N	Sialin	2.A.1.14.10	e-28	(ACS) Family
tetur533g00010	265	5	N	Sialin	2.A.1.14.10	e-32	(ACS) Family

Table IV.4 (continued)

10082							
Tetur ID[1]	length (AA)	TM[2]	S_p[3]	TCID name of best blastp hit at TCDB[4]	TCID	E-value	TCDB family
tetur01g00770	486	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-20	(FHS) Family*
tetur01g06570	444	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-13	(FHS) Family
tetur01g06580	441	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-22	(FHS) Family
tetur01g06600	441	10	N	Na+-dependent glucose transporter	2.A.1.7.4	e-26	(FHS) Family
tetur01g08660	463	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-18	(FHS) Family
tetur01g10420	457	10	N	Na+-dependent glucose transporter	2.A.1.7.4	e-21	(FHS) Family
tetur01g10430	457	10	N	Na+-dependent glucose transporter	2.A.1.7.4	e-20	(FHS) Family
tetur01g15760	460	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-24	(FHS) Family
tetur02g11610	453	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-16	(FHS) Family
tetur02g13340	467	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-09	(FHS) Family
tetur04g03830	452	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-16	(FHS) Family
tetur04g07010	458	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-21	(FHS) Family
tetur05g05100	443	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-22	(FHS) Family
tetur07g06710	443	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-19	(FHS) Family
tetur07g06730	448	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-21	(FHS) Family
tetur07g06740	447	10	N	Na+-dependent glucose transporter	2.A.1.7.4	e-24	(FHS) Family
tetur10g00280	136	2	N	Putative vanillate porter	2.A.1.15.6	0.0025	(AAHS) Family*
tetur10g00320	453	11	N	Na+-dependent glucose transporter	2.A.1.7.4	e-17	(FHS) Family
tetur13g02590	442	10	N	Na+-dependent glucose transporter	2.A.1.7.4	e-05	(FHS) Family
tetur13g02600	442	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-10	(FHS) Family
tetur13g02620	442	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-14	(FHS) Family
tetur15g00220	465	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-22	(FHS) Family
tetur18g01050	445	10	N	Na+-dependent glucose transporter	2.A.1.7.4	e-19	(FHS) Family
tetur19g02710	455	11	N	Na+-dependent glucose transporter	2.A.1.7.4	e-12	(FHS) Family
tetur21g00030	475	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-25	(FHS) Family
tetur21g00050	475	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-21	(FHS) Family
tetur21g00410	458	10	N	Na+-dependent glucose transporter	2.A.1.7.4	e-17	(FHS) Family
tetur21g00420	461	10	N	Na+-dependent glucose transporter	2.A.1.7.4	e-22	(FHS) Family
tetur21g00500	461	10	N	Na+-dependent glucose transporter	2.A.1.7.4	e-18	(FHS) Family
tetur21g00590	142	4	N	E protein viroporin	1.A.65.1.2	0.0685	E protein viroporin
tetur22g02590	449	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-13	(FHS) Family
tetur24g01860	448	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-15	(FHS) Family
tetur24g02320	449	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-09	(FHS) Family
tetur39g00620	444	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-14	(FHS) Family
tetur40g00020	484	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-37	(FHS) Family
tetur40g00030	436	12	N	Na+-dependent glucose transporter	2.A.1.7.4	e-20	(FHS) Family
10032							
Tetur ID[1]	length (AA)	TM[2]	S_p[3]	TCID name of best blastp hit at TCDB[4]	TCID	E-value	TCDB family
tetur02g07750	490	11	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0001	PCFT/HCP Family*
tetur02g07790	476	11	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0033	PCFT/HCP Family
tetur02g07920	490	11	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0002	PCFT/HCP Family
tetur02g10410	559	12	N	Quinolone resistance protein norA	2.A.1.2.10	0.0117	(DHA1) Family*
tetur03g00780	507	11	N	Tetracycline:H+ antiporter	2.A.1.2.4	0.3359	(DHA1) Family
tetur03g02680	508	10	N	Solute carrier family 46, member 3	2.A.1.50.3	e-06	PCFT/HCP Family
tetur03g02740	507	11	N	Thymic stromal cotransporter, TSCOT	2.A.1.50.2	e-05	PCFT/HCP Family
tetur03g02780	506	10	N	Thymic stromal cotransporter, TSCOT	2.A.1.50.2	0.0002	PCFT/HCP Family
tetur03g02800	512	12	N	Tetracycline:H+ antiporter	2.A.1.2.4	e-05	(DHA1) Family
tetur03g02880	508	10	N	Proton-coupled folate transporter/ Heme carrier protein	2.A.1.50.1	e-05	PCFT/HCP Family
tetur03g02890	508	10	N	Tetracycline:H+ antiporter	2.A.1.2.4	e-05	(DHA1) Family
tetur03g04410	574	12	N	Solute carrier family 46, member 3	2.A.1.50.3	e-05	PCFT/HCP Family
tetur04g02440	491	10	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0002	PCFT/HCP Family
tetur04g06740	550	12	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0001	PCFT/HCP Family
tetur06g05250	599	10	N	Tetracycline-specific exporter, TetA39 (most like 2.A.1.2.4)	2.A.1.2.38	0.0044	(DHA1) Family
tetur08g00460	496	12	N	Solute carrier family 46, member 3	2.A.1.50.3	e-05	PCFT/HCP Family
tetur08g00470	501	11	N	Proton-coupled folate transporter/ Heme carrier protein	2.A.1.50.1	e-09	PCFT/HCP Family
tetur08g00480	534	10	N	Proton-coupled folate transporter/ Heme carrier protein	2.A.1.50.1	e-05	PCFT/HCP Family
tetur08g02850	617	11	N	Thymic stromal cotransporter, TSCOT	2.A.1.50.2	e-06	PCFT/HCP Family
tetur08g04870	511	11	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0003	PCFT/HCP Family
tetur09g02400	492	11	N	Proton-coupled folate transporter/ Heme carrier protein	2.A.1.50.1	0.0502	PCFT/HCP Family
tetur09g02410	525	11	N	Glycolipid translocase (floppase) Spr1816/Spr1817	3.A.1.142.1	0.0541	(G.L.Flippase) Family*
tetur09g02420	510	10	N	Cu+ /Ag+ efflux pump	2.A.6.1.4	0.0682	(HME) Family*
tetur09g02430	525	10	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0109	PCFT/HCP Family
tetur11g00550	537	8	N	(Mono- and divalent organocation):H+ antiporter.	2.A.1.3.4	0.0017	(DHA2) Family*
tetur11g05100	501	11	N	Thymic stromal cotransporter, TSCOT	2.A.1.50.2	e-05	PCFT/HCP Family*
tetur11g05110	500	11	N	Thymic stromal cotransporter, TSCOT	2.A.1.50.2	0.0001	PCFT/HCP Family
tetur11g05410	506	10	N	Solute carrier family 46, member 3	2.A.1.50.3	e-05	PCFT/HCP Family
tetur11g06110	528	9	N	Thymic stromal cotransporter, TSCOT	2.A.1.50.2	e-06	PCFT/HCP Family
tetur13g03400	502	10	N	Solute carrier family 46, member 3	2.A.1.50.3	e-05	PCFT/HCP Family

Table IV.4 (continued)

10032							
Tetur ID[1]	length (AA)	TM[2]	S_p[3]	TCID name of best blastp hit at TCDB[4]	TCID	E-value	TCDB family
tetur14g02020	564	12	N	Thymic stromal cotransporter, TSCOT	2.A.1.50.2	0.0014	PCFT/HCP Family
tetur15g01050	489	12	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0131	PCFT/HCP Family
tetur16g01660	566	12	N	MFS permease of unknown function	2.A.1.66.1	0.007	(UMF15) Family*
tetur16g03200	499	8	N	TetA42 from a deep terrestrial subsurface bacterium	2.A.1.2.41	0.0228	(DHA1) Family
tetur16g03270	497	10	N	Solute carrier family 45, member 4	2.A.2.4.7	0.0078	GPH:CS Family
tetur17g03580	499	12	N	Thymic stromal cotransporter, TSCOT	2.A.1.50.3	e-10	PCFT/HCP Family
tetur17g00880	606	12	N	Thymic stromal cotransporter, TSCOT	2.A.1.50.2	0.0375	PCFT/HCP Family
tetur17g00890	541	12	N	Solute carrier family 45, member 4	2.A.1.50.3	0.0051	PCFT/HCP Family
tetur20g03210	501	11	N	Tetracycline:H+ antiporter	2.A.1.2.4	0.0021	(DHA1) Family
tetur30g01490	471	11	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0002	PCFT/HCP Family
tetur30g02070	549	12	N	Proton-coupled folate transporter/ Heme carrier protein	2.A.1.50.1	0.0256	PCFT/HCP Family
tetur02g07760	432	10	Y	Solute carrier family 46, member 3	2.A.1.50.3	0.0008	PCFT/HCP Family
tetur03g02670	372	7	N	Thymic stromal cotransporter, TSCOT	2.A.1.50.2	0.0002	PCFT/HCP Family
tetur03g02770	166	4	N	-	-	-	-
tetur03g02820	342	10	N	Thymic stromal cotransporter, TSCOT	2.A.1.50.2	0.0003	PCFT/HCP Family
tetur03g04360	194	11	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0036	PCFT/HCP Family
tetur03g04370	157	3	N	Tetracycline resistance protein	2.A.1.2.41	0.8993	(DHA1) Family
tetur03g04390	219	6	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0012	PCFT/HCP Family
tetur11g05550	508	11	N	Solute carrier family 46, member 3	2.A.1.50.3	0.003	PCFT/HCP Family
tetur16g03180	483	9	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0026	PCFT/HCP Family
tetur17g03650	307	4	N	Lipopolysaccharide (colanic acid) exporter, WzxC	2.A.66.2.7	0.8899	(PST) Family*
tetur46g00180	355	8	N	Inner membrane protein YqcE	2.A.1.52.2	0.3687	(UMF8) Family*
tetur623g00010	141	4	N	Solute carrier family 46, member 3	2.A.1.50.3	0.0012	PCFT/HCP Family

[1] *T. urticae* accession numbers can be accessed through the ORCAE genome portal (http://bioinformatics.psb.ugent.be/orcae/overview/Tetur)
[2] transmembrane regions were predicted using TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/)
[3] Sp : signal peptides were predicted using SignalP 3.0 (Berndtsen et al. 2004)
[4] TCID = Transporter Class ID at the Transporter Class DataBase (Saier et al. 2009)
* ACS: Anion:Cation Symporter; FHS: Fucose/H+Symporter; AAHS: Aromatic Acid:H+Symporter; PCFT/HCP: Proton-coupled folate transporter/Heme carrier protein; DHA1: Drug:H+ Antiporter-1; DHA2: Drug:H+ Antiporter-2 (14 Spanner); G.L.Flippase: Glycolipid Flippase; HME: Heavy Metal Efflux; UMF-15: Unidentified Major Facilitator-15; GPH:CS: GPH: Cation Symporter; PST: Polysaccharide Transport; UMF8:Unknown Major Facilitator-8

3.7 Transcription factors

Transcription factors belonging to the nuclear receptor family, the bHLH-PAS family, as well as to the bZIP family are known to be involved in the response to stress and xenobiotics in vertebrates and in insects (Pascussi *et al.* 2008; Misra *et al.* 2011). The *T. urticae* genome harbors at least 700 transcription factors (Grbić *et al.* 2011), and in a hierarchical clustering analysis (Pearson centered distance metric, complete linkage rule) with transcription factor expression data, Tomato-5G clustered with MR-VP and MAR-AB, and formed a sister clade to Tomato-2h and Tomato-12h. Seventeen, 20 and 27 transcription factors were differentially expressed ($|\log_2(FC)| \geq 1$, FDR < 0.05) in MR-VP, MAR-AB and Tomato-5G, respectively, although only four (*tetur03g03150*, *tetur07g01800*, *tetur24g02450* and *tetur36g00260*) were shared between MR-VP, MAR-AB and Tomato-5G (Table IV.5). *Tetur36g00260* belongs to the class of nuclear receptors and is one of the eight paralogs of the vertebrate xenosensors *PXR* and *CAR* found in the spider mites (Grbić *et al.* 2011).

In *Drosophila*, the xenosensor *CncC* is downregulated by Keap1, a Kelch-like protein (Misra *et al.* 2011). Intriguingly, we found a large proportion of genes in the OrthoMCL cluster 10254, with PFAM-domains typical for Kelch-like proteins (Table IV.1), to be differentially expressed in Tomato-5G, MR-VP and MAR-AB, with almost all (83% - 100%) of these genes downregulated.

Table IV.5 - Differentially expressed transcription factors in MR-VP, MAR-AB and Tomato-5G (Tom-5G). Transcription factors differentially expressed in both MR-VP, MAR-AB and Tomato-5G are indicated in bold.

Tetur ID[1]	OrthoMCL cluster	PFAM domain (name) ≤ e^{-5}	log_2(FC)		
			MR-VP	MAR-AB	Tom-5G
tetur08g07430	arthro11677	PF00010.21 (Helix-loop-helix DNA-binding domain), PF08447.6 (PAS fold domain)	0,00	0,00	-1,22
tetur08g07600	arthro13628	PF00010.21 (Helix-loop-helix DNA-binding domain), PF08447.6 (PAS fold domain)	0,00	-1,01	0,00
tetur03g03150	**arthro13720**	**PF00010.21 (Helix-loop-helix DNA-binding domain)**	**-1,08**	**-1,26**	**-1,67**
tetur01g11390	arthro18772	PF00010.21 (Helix-loop-helix DNA-binding domain)	1,62	0,00	1,96
tetur01g11460	arthro18772	PF00010.21 (Helix-loop-helix DNA-binding domain)	1,65	0,00	1,89
tetur01g11400	arthro18772	PF00010.21 (Helix-loop-helix DNA-binding domain)	0,00	-1,33	-1,09
tetur07g01420	SINGLETON	PF00010.21 (Helix-loop-helix DNA-binding domain)	1,61	0,00	0,00
tetur02g09640	arthro13477	PF00046.24(Homeobox)	0,00	-1,21	0,00
tetur09g01820	arthro20299	PF00046.24(Homeobox), PF00157.12 (Pou-domain)	0,00	0,00	-1,03
tetur01g13830	arthro21308	PF00046.24 (Homeobox)	0,00	0,00	-1,17
tetur24g02450	**SINGLETON**	**PF00046.24 (Homeobox)**	**1,77**	**1,29**	**1,78**
tetur21g02710	SINGLETON	PF00046.24 (Homeobox)	0,00	-2,40	0,00
tetur01g09330	arthro10817	PF00076.17 (RNA recognition motif 1)	0,00	-1,19	0,00
tetur36g00260	**arthro18765**	**PF00104.25 (Ligand-binding domain of nuclear hormone receptor)**	**1,08**	**1,80**	**1,26**
tetur03g02550	SINGLETON	PF00104.25 (Ligand-binding domain of nuclear hormone receptor)	-4,16	0,00	0,00
tetur34g00430	SINGLETON	PF00105.13 (Zinc finger, C4 type)	0,00	0,00	-1,04
tetur03g03610	arthro21359	PF00170.16 (bZIP transcription factor)	0,00	0,00	1,15
tetur02g02610	arthro13949	PF00250.13 (Fork head domain)	1,05	0,00	1,36
tetur14g02480	SINGLETON	PF00250.13 (Fork head domain)	0,00	0,00	-1,06
tetur13g04490	SINGLETON	PF00412.17 (Lim domain)	1,12	0,00	0,00
tetur18g00570	arthro17249	PF00628.24 (PHD-finger)	1,33	0,00	0,00
tetur18g00520	arthro17249	PF00628.24 (PHD-finger)	0,00	-1,21	0,00
tetur07g01800	**arthro15823**	**PF00651.26 (BTB/POZ domain)**	**1,20**	**1,08**	**1,24**
tetur11g00680	SINGLETON	PF00651.26 (BTB/POZ domain)	-1,26	0,00	0,00
tetur11g01870	SINGLETON	PF00651.26 (BTB/POZ domain)	0,00	-1,14	0,00
tetur41g00270	SINGLETON	PF00751.13 (DM DNA binding domain)	0,00	0,00	-1,02
tetur03g01600	arthro14230	PF08447.6 (PAS fold domain), PF00989.19 (PAS domain)	0,00	-1,01	0,00
tetur09g05790	SINGLETON	PF01388.16 (ARID/BRIGHT DNA binding domain)	0,00	1,07	1,02
tetur21g01360	arthro18850	PF01426.13 (Bromo adjacent homology)	0,00	1,22	0,00
tetur35g00700	SINGLETON	PF02257.10 RFX (DNA-binding domain)	0,00	0,00	-1,00
tetur08g07670	SINGLETON	PF03166.9 (MAD homology 2 domain)	0,00	1,16	1,25
tetur05g03220	arthro18519	PF05920.6 (Homeobox KN domain)	0,00	0,00	-1,17
tetur01g09160	arthro15791	PF07716.10 (Basic Leucine Zipper Domain)	0,00	-1,14	0,00
tetur27g00310	arthro13959	PF10163.4 (Transcription factor e(y)2)	0,00	-1,03	0,00
tetur12g02760	arthro10838	PF13465.1 (Zinc-finger double domain)	-1,39	0,00	0,00
tetur01g07180	arthro21295	PF13465.1 (Zinc-finger double domain)	0,00	-1,26	0,00
tetur04g03720	arthro21379	PF13465.1 (Zinc-finger double domain)	0,00	0,00	-1,81
tetur01g09970	SINGLETON	PF13465.1 (Zinc-finger double domain)	0,00	0,00	-1,03
tetur03g05420	SINGLETON	PF13551.1 (Winged helix-turn helix)	0,00	-1,40	-1,26
tetur30g00520	SINGLETON	PF13815.1 (Iguana/Dzip1-like DAZ-interacting protein N-terminal)	0,00	1,10	0,00
tetur07g07990	arthro12151	PF13873.1 (Myb/SANT-like DNA-binding domain)	0,00	0,00	-1,35
tetur34g00350	arthro12151	PF13873.1 (Myb/SANT-like DNA-binding domain)	0,00	0,00	-1,35
tetur46g00100	arthro10157	-	0,00	0,00	1,37
tetur86g00030	arthro10157	-	-1,34	0,00	-1,01
tetur01g08230	arthro10157	-	-1,29	0,00	0,00
tetur01g08270	arthro10157	-	-1,27	0,00	0,00
tetur86g00020	arthro10157	-	-1,34	0,00	0,00
tetur12g02560	arthro15835	-	0,00	0,00	-1,28

[1] *T. urticae* accession numbers can be accessed through the ORCAE genome portal (http://bioinformatics.psb.ugent.be/orcae/overview/Tetur); Tetur IDs shaded in grey are nuclear receptors according to Grbić *at al.* 2011.

Four genes were downregulated under all three conditions, in particular *tetur24g00770* (downregulated 34.8, 7.5 and 2.9 times in Tomato-5G, MR-VP and MAR-AB, respectively). We also searched for homologs of the *Drosophila* xenosensor *CncC* and identified two paralogs of this bZIP transcription factor in the *T. urticae* genome (*tetur07g06850* and *07g04600*, E-value < e^{-31} using CncC of *D. melanogaster* (GenBank accession number AAN13930) as query). However, neither changed expression levels in our microarray experiments.

3.8 Effects of host plant shift effects on acaricide toxicity

Because we found that expression profiles were highly correlated between mites adapted to tomato and pesticide-resistant strains, we asked whether acaricide toxicity was affected by host-plant shift. We chose five acaricides with four different modes of action and compared their toxicity on susceptible (London) mites kept on beans versus mites adapted to tomato. The adaptation to tomato was accompanied by a significant decrease in toxicity for three of the acaricides (Fig. IV.8). Among these, the two mitochondrial electron transport inhibitors, pyridaben and tebufenpyrad, had a different response. The toxicity of pyridaben was greatly diminished, while that of tebufenpyrad was not affected.

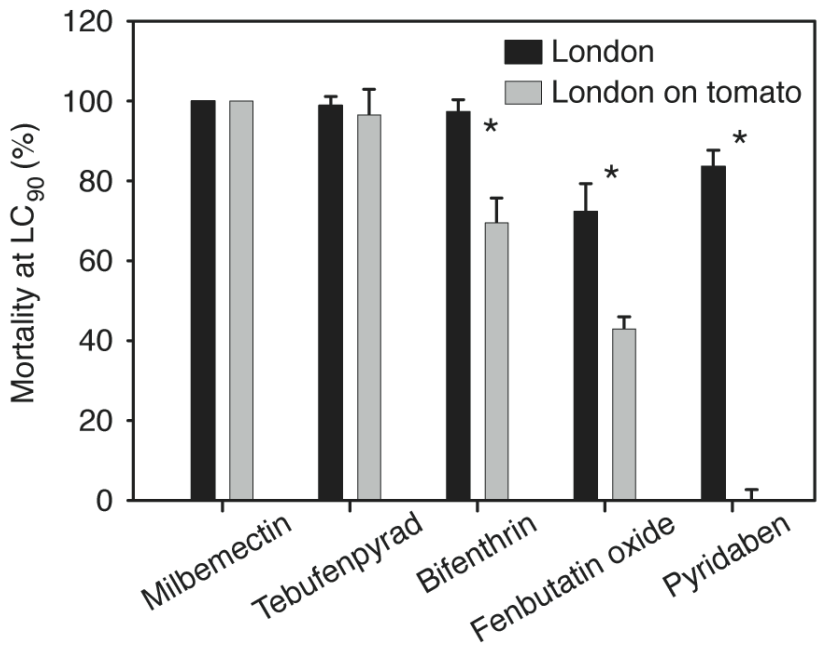


Figure IV.8 - Effect of host plant on acaricide toxicity

Percentage mortality of spider mites from the London strain to pesticide treatment. Shown are LC_{90} values for exposure to various acaricides (pyridaben, milbemectin, tebufenpyrad, fenbutatin oxide and bifenthrin) before (London) and after adaptation on tomato (London on tomato). An asterisk indicates significant differences determined by a *t*-test for paired samples.

4 Discussion

4.1 The transcriptional response to host plant shift

Classical studies of mite host transfer have shown that fitness on new hosts can increase rapidly over a small number of generations (Fry 1989; Yano *et al.* 2001). Here, we show that the response of mites to a new host is accompanied by similarly rapid changes in transcriptional profiles on a timescale of hours to a relatively small number of generations. Genes that responded in the short-term (12 hours or less) included those in classical detoxification families, such as P450s and CCEs. This finding is consistent with that observed for *T. urticae* unfed larvae that were placed for 12 hours on either bean, tomato or *Arabidopsis thaliana*, and that was ascertained with another method (RNA-seq; Grbić *et al.* 2011). However, compared to the short-term response, expression changes were far more dramatic at five generations. About 3-fold more genes were detected as differentially expressed, with ~7.5% of all genes changed in expression. The host adaptation experiments used the London reference strain that, while partially inbred, nonetheless segregates for several haplotypes throughout much of the genome (Van Leeuwen *et al.* 2012). Whether the pronounced differences in gene expression observed at five generations reflect selection on standing genetic variation, or alternatively physiological or epigenetic changes, remains to be determined. However, genetic variation between spider mites for characters affecting host use has been reported preciously (Fry 1989; Gotoh *et al.* 1993; Kant *et al.* 2008), and selection is observed in spider mite populations after a few generations (Fry 1989; Yano *et al.* 2001).

4.2 Host plant shift and pesticide resistance

Remarkably, the transcriptome of pesticide susceptible mites grown for five generations on tomato was closer to that of two acaricide-resistant strains than to that of the initial response to the host plant shift. Of note, coordinated transcriptional changes were apparent for known major environmental response genes including P450s, GSTs, and CCEs (Berenbaum 2002). This suggests that in response to different chemical challenges, the spider mite is "rounding up the usual suspects". The mechanism involved is still uncertain, but the pattern of expression for several transcription factor genes and their regulators is also similar between Tomato-5G and the resistant strains, possibly underlying the coordinated response.

In mammals, the pregnane X receptor (*PXR*) and constitutive androstane receptor (CAR) are major xenosensor genes (Pascussi *et al.* 2008). The *Drosophila* paralog of these nuclear

receptor genes, *DHR96*, is also involved in the response to xenobiotics (King-Jones *et al.* 2006). One of the eight paralogs of *DHR96*, *tetur36g00260,* was more than 2-fold upregulated in Tomato-5G, MAR-AB, and MR-VP (Table IV.5). Further, RNAi silencing of the *Drosophila Keap1* (Kelch-like ECH associated protein) gene, a negative regulator of the *Drosophila* transcription factor *CncC* (cap 'n' collar isoform-C), upregulated detoxification gene expression (Misra *et al.* 2011). In our experiments, the downregulation of OrthoMCL cluster 10254 transcripts (Kelch-like proteins) might similarly result in the upregulation of *T. urticae* detoxification genes. However, there are multiple Kelch like proteins and 2 *CncC* orthologs in *T. urticae*, and experimental studies will be required to confirm the roles of upstream transcriptional regulators in establishing the observed transcriptional profiles.

4.3 Extending the arsenal of environmental response genes

Despite the involvement of known detoxification gene families, a major finding was that, after host plant shift or between acaricide susceptible and resistant strains, many genes that encode proteins without homology to proteins of known functions changed expression. In part, this may reflect the phylogenetic distance between spider mites and insects for which most functional-genetic studies in arthropods have been performed (the divergence between mites and insects is more than 450 Myr (Dunlop 2010)). However, it may also reflect the recruitment of diverse genes to contribute to *T. urticae*'s polyphagous life style.

This interpretation is supported by our finding of strong transcriptional responses to changes in chemical exposure for some gene families either absent from insects or previously not recognized to play a major role in xenobiotic response. For instance, intradiol ring-cleavage dioxygenases, which have not been reported in other metazoa, may contribute prominently to the spider mite detoxification arsenal. The expansion of this family in the *T. urticae* genome (Grbić *et al.* 2011), and the transcriptional regulation of many family members in response to host transfer or in pesticide resistant strains, are indicative of a selective advantage after the initial horizontal transfer. Currently, the substrate specificity of the spider mite ID-RCDs is not known. Spider mite ID-RCDs are most closely related to fungal enzymes, and an ID-RCD of *Aspergillus fumigatus* can modify an array of procyanidins, which are polymers of (+)-catechine and/or (-) epicatechin (Roopesh *et al.* 2010). This hints that some ID-RCDs can metabolize more complex structures than simple catecholic substances. The *A. fumigatus* enzyme shows common features with the spider mite enzymes: it is predicted to be secreted, has a similar distance between active site residues, and contrasts with previously

biochemically characterized enzymes in bacteria and other fungi (Fig. IV.5). One of the *T. urticae* enzymes (*tetur20g01790*), has identical residues at the predicted binding site for epicatechin in the *A. fumigatus* protein sequence (Tyr184, Thr229, Arg231 and His236, *A. fumigatus* numbering) (Roopesh *et al.* 2012) (Fig. IV.5). Interestingly, a secreted ID-RCD from the oomycete *P. infestans* (GenBank accession number XP_002905783) clustered within the group of fungal secreted ID-RCDs (Fig. IV.6). This is in agreement with Richards *et al*. (2006; 2011b) who showed that the closely related species *Phytophthora ramorum* acquired its extracellular ID-RCDs from filamentous ascomycetes through horizontal gene transfer.

Other gene families are ubiquitous in their phylogenetic distribution, but their role in environmental response is striking in the spider mite. Lipocalins are small extracellular proteins and are characterized by: 1) their ability to bind to hydrophobic molecules; 2) a conserved secondary structure (an antiparallel β-barrel, with a repeated +1 topology, with an internal ligand-binding site); 3) low sequence conservation (< 20%); and 4) a conserved exon-intron structure (Flower *et al.* 2000; Sanchez *et al.* 2006). Members of the lipocalin family are found in a wide range of species, with roles in metabolism, coloration, perception, reproduction, developmental processes and modulation of immune and inflammatory responses (Chudzinski-Tavassi *et al.* 2010), resulting in a very diverse nomenclature for each specific lipocalin (e.g. apolipoprotein D, crustacyanins, biliproteins, salivacalins). As lipocalins typically bind hydrophobic molecules, they may bind pesticides/allelochemicals in mites, resulting in sequestration of these toxic, generally hydrophobic compounds. Moreover, the feeding strategy of mites may favor sequestration in dispensable phagocytes, as suggested by Mullin and Croft (Mullin and Croft 1983). Lipocalins may also protect against oxidative stress, as loss of the *Drosophila* homolog of ApoD, *Glial Lazarillo*, increased sensitivity to oxidative stress, while overexpression increased hyperoxia tolerance (Walker *et al.* 2006). Similar findings were also reported for ApoD in mice and in *Arabidopsis* (Charron *et al.* 2008; Ganfornina *et al.* 2008). In plants, the oxidative burst is one of the earliest observable aspects of the plant's defense strategy against herbivores (Hildebrand *et al.* 1986; Bi and Felton 1995). Whether mite lipocalins upregulated specifically in the early response to transfer to tomato (2 and 12h) are involved in resistance to the oxidative response warrants investigation.

In addition to enzymes and small secreted proteins, membrane binding proteins and transporters featured prominently in our analysis, including the MFS family, one of the

largest families of membrane transporters along with ABC transporters. Our experiments did not reveal an increased expression of many ABC transporter genes, although this gene family is highly expanded in the spider mite (Grbić *et al.* 2011). Multiple MFS family members were differentially regulated in both Tomato-5G and resistant strains. MFS transporters are single-polypeptide carriers that work in symport/antiport (Pao *et al.* 1998; Reddy *et al.* 2012), and studies in bacteria and fungi have identified roles in the transport of toxic substances (Hillen and Berens 1994; Saidijam *et al.* 2006; Kretschmer *et al.* 2009). For example, overexpression of the *mfsM2* gene in a sensitive strain of *Botrytis cinerea*, a fungal plant pathogen, led to drug resistance levels similar to those of a fungicide resistant *B. cinerea* strain (Kretschmer *et al.* 2009). If spider mite MFS proteins function as efflux transporters, their upregulation might result in a higher efflux of acaricides/toxic allelochemicals or their metabolites out of spider mite cells.

A third group of non-catalytic binding proteins/transporters proteins with a high percentage of genes changing expression in Tomato-5G, MR-VP, and MAR-AB were Low-Density Lipoprotein receptor proteins (LDLRs) (OrthoMCL cluster 10364, Table IV.2). In humans, these endocytic receptors bind hydrophobic molecules (Krieger *et al.* 1979) and participate in a wide range of physiological processes. Some members of this family have also multiligand binding properties (Herz and Strickland 2001; May *et al.* 2007). In our experiments, all these genes were downregulated. If hydrophobic pesticides and allelochemicals interact with these LDLRs, a downregulation could result, through lower receptor mediated endocytosis, in a lower uptake into spider mite cells.

The role of binding proteins/transporters in the response of insects to chemically challenging environments has until now been generally overlooked, and would merit closer attention. Earlier work using dedicated microarrays (Daborn *et al.* 2002; David *et al.* 2005) could not, by definition, uncover the importance of these new players. However, microarray studies with restricted random sets of ESTs already pointed out that transporters were differentially regulated in lepidopteran larvae that were fed with plants that differed in their chemical defense profile (Govind *et al.* 2010). Our study with a pangenomic array extends this early observation, and emphasizes the importance of (largely) unbiased approaches for studies to understand the basis of generalist herbivore life histories. Obtaining formal evidence that some members of these gene families actually contribute to xenobiotic tolerance when up- or down-regulated should be a priority for future research.

4.4 Host plant change and acaricide tolerance

We show here that adaptation to tomato changes not just the transcript levels of many detoxification enzymes, but results in a decreased acute toxicity for three out of five acaricides tested. The transcriptional remodeling we observed after adaptation to tomato may thus be the proximal cause for the well-known effect of host plants on the efficacy of acaricides in *T. urticae* (Neiswander *et al.* 1950; Gould *et al.* 1982; Yang *et al.* 2001; Ibrahim 2009). Host plants are also known to affect the toxicity of insecticides to insects (Brattsten *et al.* 1977; Yu *et al.* 1979; Berry *et al.* 1980; Kennedy 1984; Ahmad 1986; Lindroth 1989; Li *et al.* 2000; Li *et al.* 2004; Sasabe *et al.* 2004; Sen Zeng *et al.* 2007; Castle *et al.* 2009). Feeding on alternative hosts changes the activity of mite detoxification enzymes measured with some model substrates (Mullin and Croft 1983; Yang *et al.* 2001), and performance of mites on tomato was negatively affected by a P450 inhibitor (Agrawal *et al.* 2002). These findings indicate that the transcriptomic changes we observed may be essential for performance. The importance of active (expressed) herbivore detoxification enzymes to survival on chemically challenging hosts is also well documented in insects (e.g. (Brattsten *et al.* 1977; Snyder and Glendinning 1996)).

4.5 Polyphagy and pesticide resistance

We found a common pattern of gene expression between mites that adapted to a new host and those constitutively resistant to diverse pesticides. Moreover, the unexpected gene families that figure so prominently in our analysis, as well as the usual suspects, form a coherent whole from transcription factors to effector genes in detoxification, binding and transport. This indicates an orchestrated response rather than a random deregulation caused by the toxic effects of the plant or acaricide challenge. Many gene families we found to have strong transcriptional responses are large, reflecting a (presumed) long evolutionary history of gene duplications, sometimes in a lineage specific manner (Feyereisen 2011; Grbić *et al.* 2011). In the response to plants, such patterns are expected given the long evolutionary timescale over which plant-herbivore interactions are fine tuned. In contrast, the number of detoxification genes with altered transcription in the resistant strains, and with fold-changes mimicking that of host plant transfer, is not expected from classical theory. It has long been argued that field-evolved resistance would select single genes with major effects rather than many with limited effect (Roush and McKenzie 1987; McKenzie and Batterham 1994). In some cases of target site resistance and metabolic resistance, the monogenic inheritance has been established experimentally (ffrench-Constant *et al.* 2004; Van Leeuwen *et al.* 2012). Additionally, there

is a substantial difference between the moderate tolerance level of the tomato-adapted mites and the high resistance level of the acaricide-resistant mites. If all common genes in the transcriptomic signature contributed directly to resistance, then Tomato-5G should be much more resistant. This difference between the high resistance level of the MR-VP and MAR-AB strains and the more moderate tolerance level of the Tomato-5G mites may be due to the presence of target site mutations in the resistant strains. Indeed the presence of a mutation in a glutamate-gated chloride channel gene associated with abamectin resistance and a mutation in the sodium channel gene associated with pyrethroid resistance have already been documented for MAR-AB (see Chapter VI; John Vontas, personal communication).

We propose an explanation to these apparent paradoxes that may be relevant to the rapid development of resistance in polyphagous pests (Fig. IV.9). While a single resistance gene with major effect may eventually be selected from rare alleles, initial survival in an environment with heterogeneous dose or distribution of the pesticide may depend on multiple alleles that confer moderate resistance. Such multiple alleles may be present in modules and include genes controlling detoxification / binding / transport processes thus affecting all aspects of the toxicokinetic balance. Intraspecific genetic variation in host preference is a common aspect of polyphagy (Futuyma and Peterson 1985; Via 1990), and has been repeatedly shown in the spider mite (Gould 1979; Fry 1989). To the mite, encountering a plant treated with an acaricide may be akin to encountering a new host plant. There would be rapid selection for a genotype carrying a set of genes whose expression would best buffer against the chemical signature of the new hostile environment. That genetic variation in environmental response can come in groups of connected genes has been recently documented. Phenotypic variation in the transcriptome profile of forty inbred *Drosophila* strains was shown to consist of groups of interconnected genes. This formed hundreds of "modules" of ecologically relevant correlated genetic variation (Ayroles *et al.* 2009). In a polyphagous pest such as the spider mite, such "modules" of coregulated genes may provide an explanation for the common transcriptional patterns of MR-VP, MAR-AB and Tomato-5G. In this sense, when polyphagy is seen as genetic polymorphism in the response to different chemical environments, it may represent a pre-adaptation to xenobiotic resistance as suggested previously (Gordon 1961; Rosenheim *et al.* 1996). The selection of the rare resistance allele to the acaricide would be facilitated by the initial, higher survival rate of a subset of the population harboring it.

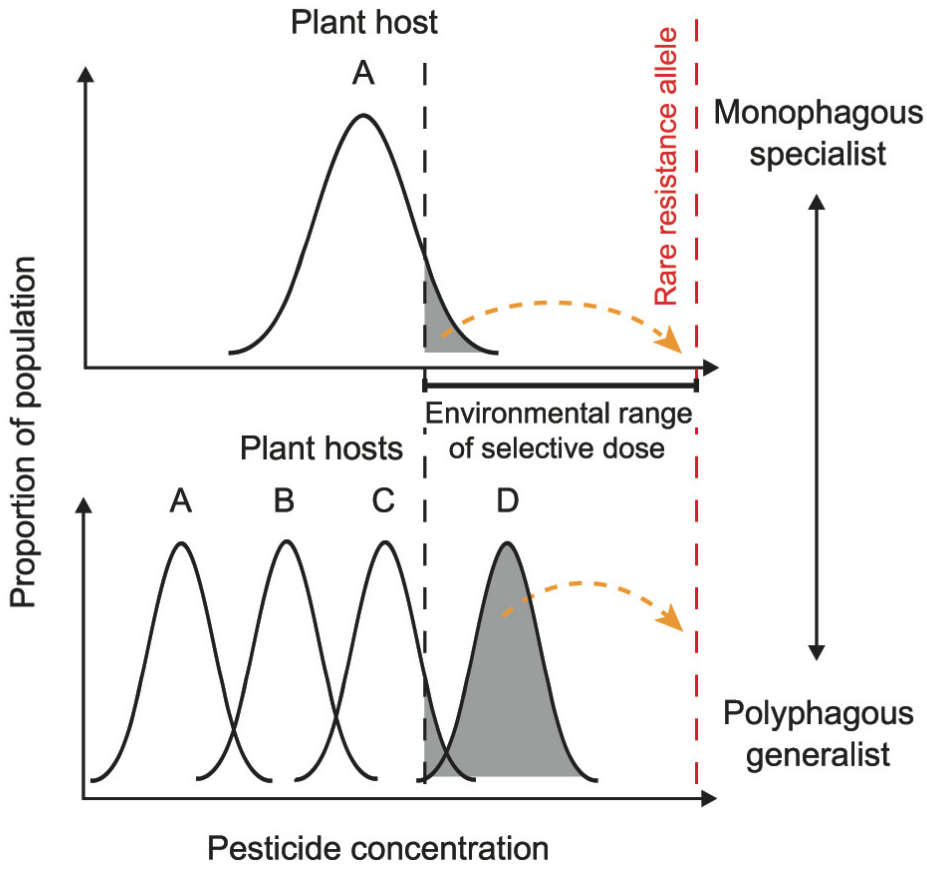


Figure IV.9 - Model for rapid evolution of pesticide resistance in generalist herbivores compared to specialists (adapted and modified from McKenzie and Batterham 1994). Preadaptation to multiple host plants is postulated to increase polymorphism in environmental responses leading to several subsets of alleles. The initial stages of selection by a pesticide mimics a host plant shift, resulting in rapid selection of the best adapted subset of environmental response alleles. This in turn provides a larger initial population from which a rare (high) resistance allele can be selected, thus accelerating the development of agriculturally significant resistance levels when compared to specialists. After both types of selection (host shift, pesticide), the transcriptome signature is similar, because it is drawn from a similar subset of genotypes.

A plant-specific transcriptional response has been observed before in polyphagous Lepidoptera and in the spider mite (Govind *et al.* 2010; Grbić *et al.* 2011; Celorio-Mancera *et al.* 2012). Specialist herbivores, on the other hand, are characterized by a xenobiotic response that is more constitutive and more targeted to the favorite host plants (Berenbaum 2002), and the transcriptional response to change in plant chemistry is much more restricted in a specialist than in a generalist (Govind *et al.* 2010). It remains to be shown that intraspecific genetic variation in the spider mite includes the differential regulation of specific subsets of genes involved in fitness on different host plants, such as detoxification, binding and

transport as we predict here. Such experimental verification is as important as it is difficult to obtain. However, the rapid development of resistance in polyphagous herbivores and its relative absence in specialists such as natural enemies and predators is well known (Croft and Strickler 1983). Our results provide an unprecedented insight into the transcriptional correlation that may link polyphagy with development of pesticide resistance. They also highlight the need to study not just detoxification enzymes, but also binding proteins and transporters as major contributors to survival in a toxic environment. The spider mite is a polyphagous jack of all trades, but has evolved to be a master of some as well.

Chapter V

A burst of ABC genes in the genome of the polyphagous spider mite *Tetranychus urticae*

This chapter has been redrafted from:

Dermauw W., Osborne E.J., Clark R.M., Grbic M., Tirry L., Van Leeuwen T. (2013). A burst of ABC genes in the genome of the polyphagous spider mite *Tetranychus urticae*. *BMC Genomics*, in press.

1 Introduction

ATP-binding cassette (ABC) proteins form one of the largest protein families that are present in all living organisms on earth. The majority of ABC proteins are membrane bound primary transporters, using ATP to translocate substrates across extra- and intracellular membranes. In addition, these ABC transporters are mostly uniporters, mediating the unidirectional translocation of a substrate (Higgins 1992; Dean *et al.* 2001; Hollenstein *et al.* 2007; Rees *et al.* 2009). The Major Facilitator Superfamily (MFS) is another large transporter family present in all living organisms, but as opposed to ABC transporters, it comprises secondary carriers that can be either uniporters, symporters or antiporters (Pao *et al.* 1998). In most ABC proteins two types of domains can be distinguished, an ATP-binding domain (also named the nucleotide binding domain (NBD)) and a transmembrane domain (TMD). The highly conserved NBD contains three motifs: a Walker A and Walker B domain and the ABC signature (LSSG-motif). The NBD binds and hydrolyses ATP and provides energy to transport substrates. The TMD consists of five to six membrane spanning helices and provides the specificity for the substrate. Full transporters comprise two NBDs and two TMDs while half transporters have only one of each type and require homo- or heterodimerization to form a functional unit (Higgins 1992; Dean *et al.* 2001; Hollenstein *et al.* 2007; Rees *et al.* 2009). Based on the homology of their NBDs, ABC proteins have been divided into seven subfamilies, ABCA to ABCH (Dean *et al.* 2001; Dean and Annilo 2005). Interestingly, the ABCH subfamily was discovered during analysis of the *Drosophila melanogaster* genome and is present in all sequenced arthropod genomes to date and teleost fish, but not in mammals, plants or fungi (Roth *et al*. 2003; Dean and Annilo 2005; Annilo *et al*. 2006; Sturm *et al*. 2009; Kovalchuk and Driessen 2010; Popovic *et al*. 2010; Labbé *et al*. 2011; Liu *et al*. 2011b; Xie *et al*. 2012; Broehan *et al*. 2013; You *et al*. 2013).

In humans, ABC proteins mainly function in the membrane transport of substrates, including amino acids, sugars, lipids, inorganic ions, polysaccharides, metals, peptides, toxic metabolites and drugs (Higgins 1992; Rees *et al.* 2009). In addition to transporters, the human ABC protein superfamily also contains ion channels (CFTR), receptors (SUR1 and 2) and proteins involved in translation (human ABCE and ABCF1, 2 and 3) (Dean *et al.* 2001). Mutations in ABC genes have been linked to several human disorders, like cystic fibrosis, adrenoleukodystrophy, sitosterolemia and diabetes (Borst and Elferink 2002; Tarr *et al.*

2009). Furthermore, within the ABCB, C and G subfamilies, many genes code for proteins that contribute to resistance of cancer cells against chemotherapeutic agents: the multidrug resistance proteins or P-glycoproteins (MDR or P-gps, members of the ABCB subfamily), the multidrug resistance-associated proteins (MRP, members of the ABCC subfamily) and the breast cancer protein (BCRP or human ABCG2) (Schinkel and Jonker 2003; Fukuda and Schuetz 2012). In insects, it has been shown that ABC transporters have functions that affect metabolism, development and resistance to xenobiotics including insecticides and plant secondary toxic compounds (allelochemicals) (Labbé *et al.* 2011). Some ABC proteins have specific functions that are well documented in arthropods. In *D. melanogaster* the ABCC transporter Mdr49 controls the export of a germ cell attractant (Ricardo and Lehmann 2009). *D. melanogaster* white, on the other hand, is a member of the ABCG subfamily and is involved in the uptake of pigment precursors in the developing eye (Mackenzie *et al.* 1999). Its ortholog in *Bombyx mori* (Bmwh3) and *T. castaneum* (TcABCG-9B) have similar functions, and *w-3*^{oe} *B. mori* mutants and *TcABCG-9B* dsRNA injected adult beetles have white eyes (Komoto *et al.* 2009; Broehan *et al.* 2013). In the tobacco hornworm, *Manduca sexta*, orthologs of human P-gps are essential as they prevent the influx of nicotine across the blood brain barrier (Murray *et al.* 1994; Gaertner *et al.* 1998). Insect orthologs of human P-gps and MRPs have also been frequently linked to pesticide resistance (Lanning *et al.* 1996; Buss and Callaghan 2008; Baxter *et al.* 2011; Heckel 2012). Resistance to pesticides in insects is either related to reduced target-site sensitivity or sequestration/metabolism of the pesticide before it reaches the target site by quantitative or qualitative changes of genes involved in the detoxification process (Li *et al*. 2007; Van Leeuwen *et al*. 2010a). These "detoxification" genes comprise members of the P450 mono-oxygenases (P450s), glutathione-S-transferases (GSTs), carboxyl/cholinesterases (CCEs) and also the less known ABC transporters. Although the ABC transporters have often been overlooked in studies that describe the detoxification toolkit in sequenced insect genomes (Claudianos *et al.* 2006; Oakeshott *et al.* 2010; Ramsey *et al.* 2010), clear examples of their importance in detoxification have been documented. For example, Lanning *et al*. (1996) showed that increased expression of human P-gp orthologues in *Heliothis virescens* was associated with resistance to thiodicarb, and a mutation in the same ABCC member of four different lepidopteran species was recently associated with resistance to the Cry1A toxin (Baxter *et al.* 2011; Heckel 2012).

Complete and correctly annotated gene inventories are a prerequisite to study the biological role and evolutionary history of ABC genes. Among arthropods, detailed studies of ABC families have been published for members of several different insect orders (Dean *et al*. 2001; Roth *et al*. 2003; Labbé *et al*. 2011; Liu *et al*. 2011b; Xie *et al*. 2012; Broehan *et al*. 2013) and the crustacean *Daphnia pulex* (Sturm *et al.* 2009). In contrast, besides the identification of nine ABC genes in the mange mite *Sarcoptes scabiei* (Mounsey *et al.* 2006), there are no reported studies in the subphylum Chelicerata (spiders, scorpions, mites and ticks), one of the most diverse groups of terrestrial animals (Walter and Proctor 1999). Recently, the first draft genome sequence of a chelicerate, the two-spotted spider mite, *Tetranychus urticae*, was reported (Grbić *et al.* 2011). The spider mite is one of the most polyphagous herbivores known, and has been documented to feed on more than 1,100 plant species that belong to more than 140 different plant families, including many that produce toxic compounds (Jeppson *et al.* 1975; Migeon and Dorkeld 2010). In addition, spider mites are major agricultural pests and are the "resistance champion" among arthropods as they have the most documented instances of resistance to diverse pesticides (Van Leeuwen *et al*. 2010a; Whalon *et al*. 2013). The molecular mechanisms underlying spider mite resistance to xenobiotics (pesticides and allelochemicals) are less understood compared to insects (Yang *et al*. 2002b; Van Leeuwen *et al*. 2010a). However, the availability of the genome sequence now provides unique information and tools for the study of the role of gene families involved in xenobiotic metabolism in spider mites (see Chapter IV) (Van Leeuwen *et al.* 2012; Van Leeuwen *et al.* 2013). Characterization of spider mite gene families associated with detoxification of xenobiotics, including ABC genes, is the first step towards a better understanding of how spider mites cope with these compounds (Van Leeuwen *et al.* 2013). An initial preliminary analysis of ABC genes in the spider mite genome focused solely on ABCB and ABCC subfamilies (Grbić *et al.* 2011), and lacked a complete description of the phylogenetic relationships with other metazoan ABCs. In this study, we provide a detailed comparison of all ABC subfamilies (ABCA-ABCH) in *T. urticae* with those of the insect *D. melanogaster*, the crustacean *D. pulex*, the nematode *Caenorhabditis elegans* and the mammal *Homo sapiens*. Further, we show that expression levels of ABC genes change in pesticide resistant strains and when new and challenging plant host are encountered. Our analyses will facilitate biochemical and toxicological studies of the role of *T. urticae* ABC transporters in spider mite physiology, and in particular the extraordinary host range and pesticide resistance development.

2 Materials and methods

2.1 Annotation and phylogeny of ABC transporters

ABCs were identified in a similar way as for *D. pulex* (Sturm *et al.* 2009). Briefly, tBLASTn searches (Altschul *et al.* 1997) were performed on the *T. urticae* genome sequence assembly (version July 2012, available at http://bioinformatics.psb.ugent.be/orcae/overview/Tetur.) using the highly conserved nucleotide binding domain (NBD) of *D. melanogaster* ABC proteins as queries. One search was carried out per subfamily, using the sequence of the NBD of a representative *D. melanogaster* protein (A: CG1718; B: CG3879 (Mdr49); C: CG9270; D: CG12703; E: CG5651; F: CG9330; G: white; H: CG9990). If the *D. melanogaster* transporter had two NBDs, the N-terminal domain was used. All hits with an E-value less than e^{-4} were withdrawn for analysis and gene models were refined or created on the basis of homology and RNA-seq support (http://bioinformatics.psb.ugent.be/orcae/overview/Tetur). The NBDs from those *T. urticae* gene models encoding complete ABCs (i.e. not lacking one or both vital domains (TMD and NBD) were extracted using the ScanProsite facility (http://prosite.expasy.org/scanprosite/) and the Prosite profile PS50893. *T. urticae* ABC protein NBDs were aligned with NBDs of *D. melanogaster* and human ABC transporters using MUSCLE (Edgar 2004). Model selection was done with Prottest 2.4 (Abascal *et al.* 2005). According to the Akaike information criterion LG+F+G was optimal for phylogenetic analysis (see Appendix V-A and V-B for protein alignment and likelihood scores, respectively). A maximum likelihood phylogenetic analysis of *T. urticae*, *D. melanogaster* and human ABC protein NBDs, bootstrapping with 1000 pseudoreplicates, was performed using Treefinder (Jobb *et al.* 2004) to confirm the position of *T. urticae* ABCs within ABC classes (A-G). A similar phylogenetic analysis, restricted to N-terminal NBDs of *T. urticae*, was also performed (optimal model according to the Akaike Information Criterion: LG+I+G+F, see Appendix V-A and V-B for protein alignment and likelihood scores, respectively). Similar to previous studies, in the phylogenetic analysis using *T. urticae*, *D. melanogaster* and human ABC protein NBDs C-terminal NBDs of the ABCC subfamily clustered together with NBDs of the ABCB subfamily (Allikmets *et al*. 1996; Sheps *et al*. 2004; Liu *et al*. 2011b). The subfamily assignment was further confirmed by BLASTp analyses of the manually corrected models on the NCBI website. We adopted the guidelines set forth by the human genome organization nomenclature committee (HGNC) for naming the *T. urticae* ABC proteins. Separate phylogenetic analyses on full ABC protein sequences

of *T. urticae*, *D. pulex*, *C. elegans*, *D. melanogaster* and *H. sapiens* ABCs were also carried out for each subfamily, using the same methodology as above. The optimal model used for maximum likelihood phylogenetic analysis of the ABCA, -B (full and half transporters), -C, -D, -E, -F, -G and –H subfamilies was LG+I+G+F (see Appendix V-A and V-B for protein alignment and likelihood scores, respectively). According to previous studies (Annilo *et al.* 2006; Sturm *et al.* 2009), a separate phylogenetic analysis for each ABC subfamily facilitates bioinformatics analyses and results in a more meaningful degree of resolution in phylogenetic analysis. Similar to other ABC-subfamily studies (Anjard *et al*. 2002, Sturm *et al.* 2009) resulting ABC family trees were midpoint rooted, a valid rooting method in the absence of a suitable outgroup (Hess and De Moraes Russo 2007). Finally, in order to detect ABC pseudogenes/fragments not containing ABC NBDs, all protein sequences of complete ABCs were used as query in tBLASTn-searches against the *T.urticae* genome. Phylogenetic trees were visualized and edited using MEGA5 (Tamura *et al.* 2011) and CorelDraw X3 (Corel Inc.), respectively.

2.2 Sequence similarity, transmembrane prediction and gene structure of *T. urticae* ABC proteins

ABC protein sequence similarities and identities were calculated using MatGAT 2.03 (Campanella *et al.* 2003) using default settings (BLOSUM50 matrix, gap opening and extending gap penalty set to 12 and 2, respectively). Transmembrane domains of *T. urticae* ABCs were predicted using the SCAMPI prediction server (Bernsel *et al.* 2008) . Subcellular localization was predicted using TargetP 1.0 (Emanuelsson *et al.* 2007) . Gene structures of *T. urticae* ABCs were visualized using the coordinates of each *T. urticae* ABC transporter (available at http://bioinformatics.psb.ugent.be/orcae/overview/Tetur) and the fancyGene visualization software (Rambaldi and Ciccarelli 2009). N-glycosylation sites were predicted with NetNGlyc1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/). Only N-glycosylation sites with a "potential" score > 0.05 and with a jury agreement ("++"-sign or higher) were included in analyses. O-glycosylation sites were predicted using NetOGlyc 3.1 server (Julenius *et al.* 2005). If the G-score was higher than 0.5 the residue was considered to be O-glycosylated. The number of O-glycoslated sites (glycosylated serines and threonines) is shown in Table V.1.

2.3 Expression profiling of ABC genes

Expression profiling of ABC genes was assessed using microarray expression data of two multi-pesticide resistant strains (MR-VP and MAR-AB) (see Chapter IV) and a previously published RNA-seq dataset (Grbić *et al.* 2011). The RNA-seq dataset consists of replicated RNA-seq libraries of spider mites feeding on different host plants (bean, tomato and *Arabidopsis*) and a single RNA-seq library for different developmental stages of spider mites (embryo, larvae, nymph and adult). Experimental details can be found in Grbić *et al*. (2011) and the RNA-seq data are available via Gene Expression Omnibus under reference GSE32342. To ensure the best possible alignment of RNA-seq reads to our manually curated ABC transporter gene models, we re-mapped the RNA-seq reads to the spider mite genome with an updated annotation (Nov 15, 2012 release; http://bioinformatics.psb.ugent.be/orcae/overview/Tetur). Read alignments and expression quantification were performed after Grbić *et al*. (2011). For host transfer experiments, differential gene expression was assessed with the DESeq R package (Anders and Huber 2010) as previously described (Grbić *et al.* 2011). For the microarray experiment, differentially expressed genes were assessed as reported earlier (see Chapter IV). For both the host transfer experiment (RNA-seq) and expression profiling with multi-pesticide resistant strains (microarrays), ABC genes with a fold change higher than two and a FDR adjusted p-value less than 0.05 were considered as differentially expressed.

3 Results and discussion

3.1 Identification of spider mite ABC genes

We identified 103 putative ABC genes in the genome of *T. urticae* (Table V.1). To our knowledge, this is the largest number of ABC genes reported for any metazoan species so far (Sheps *et al*. 2004; Liu *et al*. 2011b; Broehan *et al*. 2013; You *et al*. 2013). Of all organisms sequenced to date, only the protozoan ciliate *Tetrahymena thermophila* has more ABC genes (Xiong *et al.* 2010). A maximum likelihood phylogenetic analysis grouped the *T. urticae* ABC proteins into each of the eight known ABC subfamilies with high bootstrap support (Fig. V.1 and V.2). We identified 9, 4, 39, 2, 1, 3, 23 and 22 ABC transporters belonging to the ABCA, ABCB, ABCC, ABCD, ABCE, ABCF, ABCG and ABCH subfamilies, respectively (Table V.2). Significant homology (E-value $\leq e^{-4}$) with one of the 103 putative *T. urticae* ABC genes was found at an additional 41 loci in the *T. urticae* genome. However, gene models at these loci, most of which had homology to the ABCC or ABCH subfamilies, lacked one or both vital domains (NBDs, TMD) of canonical ABC genes. These likely represent gene fragments or pseudogenized genes, and were excluded from detailed analysis (Table V.3). From the 103 full-length *T. urticae* ABC genes, almost half (48) are located on only 5 genomic scaffolds (11, 9, 6, 11 and 10 ABC genes on scaffold 1, 3, 4, 9 and 11, respectively). In addition, a high complexity was observed within gene structures, with exon numbers ranging from 1 to 20 (Table V.1, Appendix V-E). To determine the accurate evolutionary position of the 103 *T. urticae* ABCs, a phylogenetic analysis including full-length ABC protein sequences from the draft genomes of *T. urticae* and *D. pulex* and from the finished genomes of *C. elegans*, *D. melanogaster* and *H. sapiens,* were performed for each subfamily separately. The results of these analyses are discussed in the following sections.

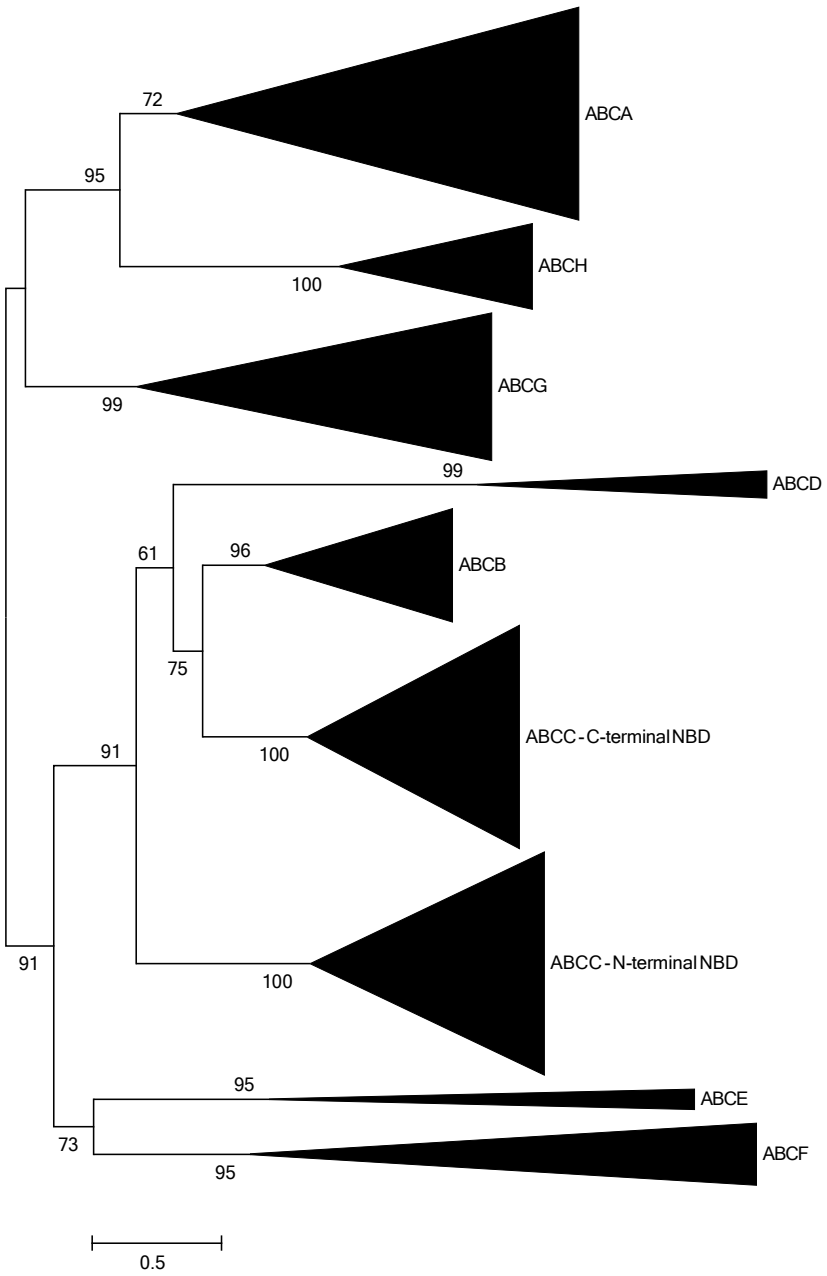


Figure V.1 - Midpoint rooted maximum likelihood phylogenetic tree of ABC protein NBDs of *D. melanogaster*, *H. sapiens* and *T. urticae*.

For amino acid alignment, amino acid substitution model and likelihood score of the constructed phylogenetic tree see Appendix V-A and V-B. Main nodes were collapsed to create a better overview of the phylogenetic relationships between the different ABC subfamilies. Numbers at the branch point of each node represent the bootstrap value resulting from 1000 pseudoreplicates (LR-ELW). The scale bar represents 0.5 amino-acid substitutions per site. For accession numbers of metazoan ABC protein sequences see Appendix V-C while *T. urticae* ABC proteins can be found in Appendix V-D.

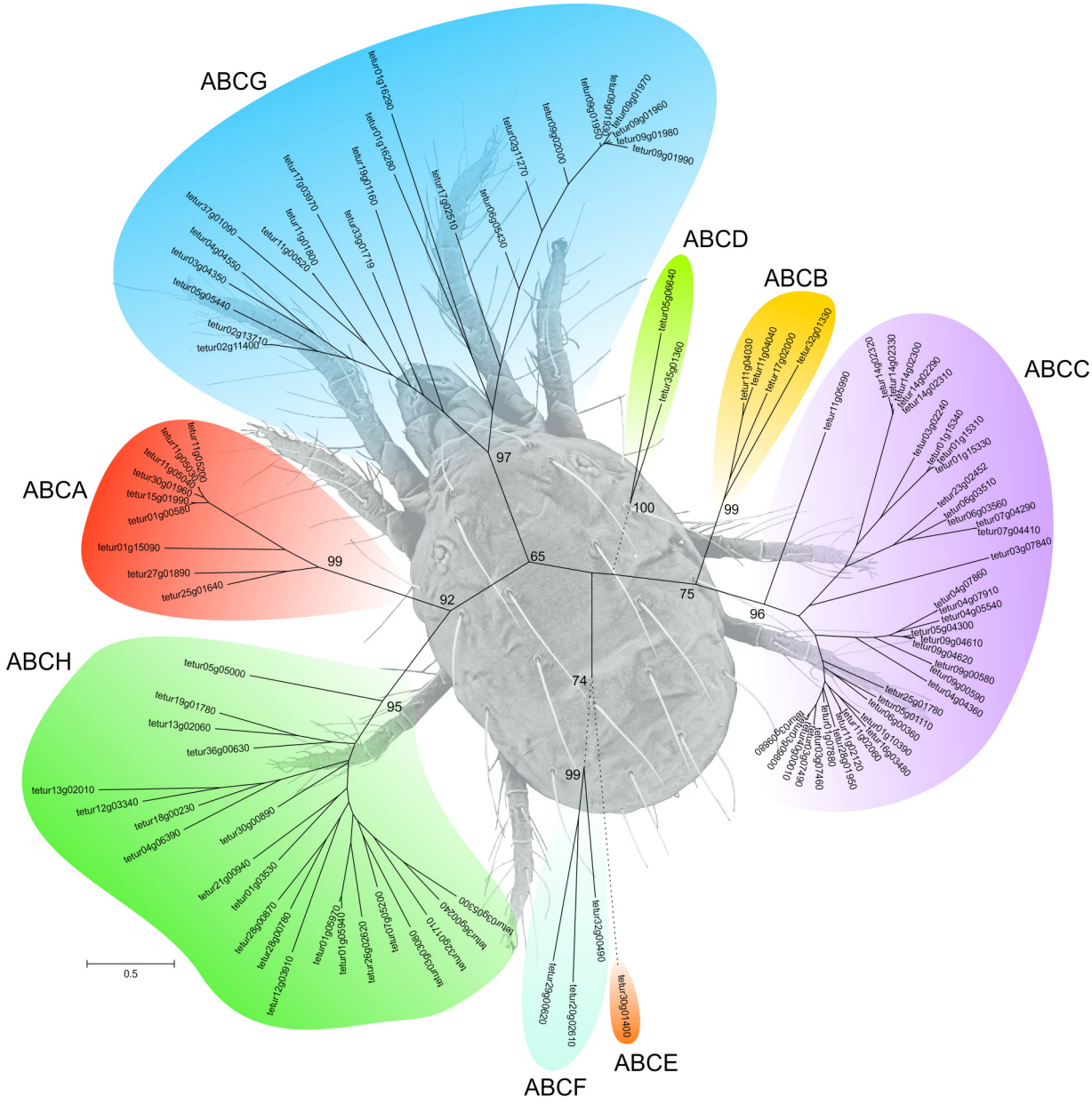


Figure V.2 – Unrooted phylogenetic tree of N-terminal NBDs of 103 ABC proteins of *T. urticae*. Amino acid sequences of NBDs were aligned using MUSCLE (Edgar 2004) and subjected to a maximum likelihood analysis using Treefinder (Jobb *et al.* 2004). For amino acid alignment, amino acid substitution model and likelihood score of the constructed phylogenetic tree see Appendix V-A and V-B. Numbers at important nodes represent the bootstrap values resulting from 1000 pseudoreplicates (LR-ELW). The scale bar represents 0.5 amino-acid substitutions per site. The different ABC transporter subfamilies are indicated by shaded colors. Accession numbers of metazoan ABC protein sequences can be found in Appendix V-C while *T. urticae* ABC protein sequences can be found in Appendix V-D.

Table V.1 - Characterisation of 103 *T. urticae* ABC proteins

Subfamily	Tetur ID	Name	Length (AA)	Strand	Exons	Topology[1]	N-Glc[2]	O-Glc[3]
A (9)	tetur01g00580	TuABCA-01	1,659	-	5	$(6TM\text{-}NBD)_2$	5	
	tetur01g15090	TuABCA-02	1,639	+	4	$(7/8TM\text{-}NBD)_2$	2	1
	tetur11g05030	TuABCA-03	1,670	-	5	$(5/6TM\text{-}NBD)_2$	3	
	tetur11g05040	TuABCA-04	1,670	-	5	$(6TM\text{-}NBD)_2$	2	
	tetur11g05200	TuABCA-05	1,672	-	5	$(6TM\text{-}NBD)_2$	2	
	tetur15g01990	TuABCA-06	1,666	+	5	$(6TM\ \text{-}NBD)_2$	3	
	tetur25g01640	TuABCA-07	2,302	+	20	$(6TM\text{-}NBD)_2$	8	
	tetur27g01890	TuABCA-08	2,082	+	2	$(7/4TM\text{-}NBD)_2$	5	27
	tetur30g01960	TuABCA-09	1,651	-	5	$(6/8TM\text{-}NBD)_2$	3	
B (4)	tetur11g04030	TuABCB-01	1,294	+	15	$(6TM\text{-}NBD)_2$	4	
	tetur11g04040	TuABCB-02	1,292	+	16	$(6TM\text{-}NBD)_2$	4	
	tetur17g02000	TuABCB-03	688	-	8	4TM-NBD	1	
	tetur32g01330	TuABCB-04	656	-	8	5TM-NBD	3	
C (39)	tetur01g07880	TuABCC-01	1,507	+	11	$5TM\text{-}(5TM\text{-}NBD)_2$	1	2
	tetur01g10390	TuABCC-02	1,529	+	11	$7TM\text{-}(5TM\text{-}NBD)_2$		
	tetur01g15310	TuABCC-03	1,299	+	15	$(5/6TM\text{-}NBD)_2$	5	1
	tetur01g15330	TuABCC-04	1,328	+	14	$(7TM\text{-}NBD)_2$	5	
	tetur01g15340	TuABCC-05	1,338	+	14	$(8/7TM\text{-}NBD)_2$	3	
	tetur03g02240	TuABCC-06	1,375	-	12	$(6TM\text{-}NBD)_2$	2	
	tetur03g07460	TuABCC-07	1,481	-	11	$6TM\text{-}(5TM\text{-}NBD)_2$	2	
	tetur03g07490	TuABCC-08	1,483	-	11	$6TM\text{-}(5TM\text{-}NBD)_2$	1	
	tetur03g07840	TuABCC-09	1,457	-	5	$5TM\text{-}(6TM\text{-}NBD)_2$	1	
	tetur03g09800	TuABCC-10	1,490	+	11	$6TM\text{-}(5TM\text{-}NBD)_2$		1
	tetur03g09880	TuABCC-11	1,495	+	11	$4TM\text{-}(5TM\text{-}NBD)_2$	1	1
	tetur04g04360	TuABCC-12	1,522	+	5	$8TM\text{-}(5TM\text{-}NBD)_2$	3	
	tetur04g05540	TuABCC-13	1,503	+	8	$6TM\text{-}(5TM\text{-}NBD)_2$	5	
	tetur04g07860	TuABCC-14	1,501	-	8	$7TM\text{-}(4TM\text{-}NBD)_2$	3	
	tetur04g07910	TuABCC-15	1,493	-	8	$6TM\text{-}(5TM\text{-}NBD)_2$	2	
	tetur05g01110	TuABCC-16	1,522	-	11	$4TM\text{-}(5TM\text{-}NBD)_2$	6	1
	tetur05g04300	TuABCC-17	1,490	+	8	$5TM\text{-}(6TM\text{-}NBD)_2$	6	2
	tetur06g00360	TuABCC-18	1,516	+	11	$6TM\text{-}(5TM\text{-}NBD)_2$	4	1
	tetur06g03510	TuABCC-19	1,343	+	6	$(6TM\text{-}NBD)_2$	2	
	tetur06g03560	TuABCC-20	1,339	+	6	$(5/7TM\text{-}NBD)_2$	2	
	tetur07g04290	TuABCC-21	1,324	-	5	$(6TM\text{-}NBD)_2$	1	
	tetur07g04410	TuABCC-22	1,324	+	5	$(6TM\text{-}NBD)_2$	1	
	tetur09g00580	TuABCC-23	1,506	+	8	$7TM\text{-}(6TM\text{-}NBD)_2$	2	
	tetur09g00590	TuABCC-24	1,496	+	9	$6TM\text{-}(6TM\text{-}NBD)_2$	1	
	tetur09g04610	TuABCC-25	1,521	+	8	$7TM\text{-}(6TM\text{-}NBD)_2$	4	1
	tetur09g04620	TuABCC-26	1,501	+	8	$6TM\text{-}(5TM\text{-}NBD)_2$	5	
	tetur11g02060	TuABCC-27	1,499	-	11	$4TM\text{-}(6TM\text{-}NBD)_2$	1	
	tetur11g02120	TuABCC-28	1,492	+	11	$6TM\text{-}(4TM\text{-}NBD)_2$	2	1
	tetur11g05990	TuABCC-29	1,683	-	15	$5TM\text{-}(6TM\text{-}NBD)_2$	6	
	tetur14g02290	TuABCC-30	1,309	-	16	$(5/7TM\text{-}NBD)_2$	4	
	tetur14g02300	TuABCC-31	1,312	-	16	$(5/7TM\text{-}NBD)_2$	5	
	tetur14g02310	TuABCC-32	1,308	-	16	$(6/7TM\text{-}NBD)_2$	3	1
	tetur14g02320	TuABCC-33	1,309	-	16	$(7TM\text{-}NBD)_2$	2	1
	tetur14g02330	TuABCC-34	1,309	-	16	$(5/7TM\text{-}NBD)_2$	2	
	tetur16g03480	TuABCC-35	1,513	-	11	$6TM\text{-}(5TM\text{-}NBD)_2$	2	1
	tetur23g02452	TuABCC-36	1,344	-	6	$(8/6TM\text{-}NBD)_2$	2	
	tetur25g01780	TuABCC-37	1,525	+	12	$5TM\text{-}(5TM\text{-}NBD)_2$		
	tetur28g01950	TuABCC-38	1,287	+	7	$(6TM\text{-}NBD)_2$	3	
	tetur40g00010	TuABCC-39	1,490	+	11	$5TM\text{-}(5TM\text{-}NBD)_2$		1

Table V.1 (continued)

Subfamily	Tetur ID	Name	Length (AA)	Strand	Exons	Topology[1]	N-Glc[2]	O-Glc[3]
D (2)	tetur05g06640	TuABCD-01	829	+	4	5TM-NBD	2	1
	tetur35g01360	TuABCD-02	656	+	5	5TM-NBD	4	
E (1)	tetur30g01400	TuABCE-01	614	-	2	NBD-NBD		
F (3)	tetur20g02610	TuABCF-01	612	+	4	NBD-NBD	3	
	tetur29g00620	TuABCF-02	584	+	1	NBD-NBD	3	
	tetur32g00490	TuABCF-03	718	-	4	NBD-NBD	4	
G (23)	tetur01g16280	TuABCG-01	915	-	6	NBD-7TM	1	1
	tetur01g16290	TuABCG-02	643	+	3	NBD-6TM	1	
	tetur02g11270	TuABCG-03	689	+	11	NBD-6TM	1	
	tetur02g11400	TuABCG-04	811	-	2	NBD-6TM	3	1
	tetur02g13710	TuABCG-05	810	-	2	NBD-6TM	3	
	tetur03g04350	TuABCG-06	687	-	2	NBD-6TM	2	
	tetur04g04550	TuABCG-07	720	+	2	NBD-6TM	3	1
	tetur05g05440	TuABCG-08	721	+	2	NBD-8TM	1	1
	tetur06g05430	TuABCG-09	646	-	12	NBD-7TM	3	
	tetur09g01930	TuABCG-10	685	+	9	NBD-5TM	2	
	tetur09g01950	TuABCG-11	682	+	9	NBD-5TM	1	
	tetur09g01960	TuABCG-12	683	+	9	NBD-5TM	3	
	tetur09g01970	TuABCG-13	687	+	9	NBD-6TM	3	
	tetur09g01980	TuABCG-14	680	+	9	NBD-6TM	4	
	tetur09g01990	TuABCG-15	687	+	9	NBD-6TM	2	4
	tetur09g02000	TuABCG-16	682	+	9	NBD-7TM	3	
	tetur11g00520	TuABCG-17	792	-	1	NBD-8TM	2	3
	tetur11g01800	TuABCG-18	790	-	1	NBD-8TM	2	3
	tetur17g02510	TuABCG-19	832	-	4	NBD-4TM	2	1
	tetur17g03970	TuABCG-20	700	-	2	NBD-8TM	5	
	tetur19g01160	TuABCG-21	775	-	1	NBD-7TM	4	2
	tetur33g01719	TuABCG-22	760	+	1	NBD-6TM	1	
	tetur37g01090	TuABCG-23	779	-	1	NBD-7TM	1	14
H (22)	tetur01g03530	TuABCH-01	827	+	1	NBD-8TM	6	7
	tetur01g05940	TuABCH-02	766	-	6	NBD-6TM	2	2
	tetur01g05970	TuABCH-03	734	-	6	NBD-6TM	1	1
	tetur03g03080	TuABCH-04	767	-	2	NBD-8TM	3	
	tetur03g05300	TuABCH-05	738	-	1	NBD-6TM	1	
	tetur04g06390	TuABCH-06	707	+	1	NBD-7TM	2	1
	tetur05g05000	TuABCH-07	713	+	1	NBD-7TM	3	
	tetur07g05200	TuABCH-08	764	-	4	NBD-6TM		
	tetur12g03340	TuABCH-09	744	-	1	NBD-5TM	1	
	tetur12g03910	TuABCH-10	713	+	1	NBD-7TM	3	
	tetur13g02010	TuABCH-11	703	-	1	NBD-8TM	2	1
	tetur13g02060	TuABCH-12	723	-	1	NBD-7TM	3	
	tetur18g00230	TuABCH-13	735	-	2	NBD-8TM	4	20
	tetur19g01780	TuABCH-14	692	-	1	NBD-7TM	5	2
	tetur21g00940	TuABCH-15	716	+	2	NBD-6TM	1	1
	tetur26g02620	TuABCH-16	708	-	10	NBD-7TM	2	4
	tetur28g00780	TuABCH-17	726	-	1	NBD-8TM	1	2
	tetur28g00870	TuABCH-18	720	+	1	NBD-7TM	3	1
	tetur30g00890	TuABCH-19	721	-	1	NBD-7TM	2	
	tetur32g01710	TuABCH-20	804	-	1	NBD-6TM	2	4
	tetur36g00240	TuABCH-21	777	-	1	NBD-6TM	3	1
	tetur36g00630	TuABCH-22	704	-	1	NBD-6TM	1	

[1] transmembrane helices (TMs) were predicted using the SCAMPI server (Bernsel *et al.* 2008) while nucleotide binding domains (NBD) were determined using the ScanProSite facility (http://prosite.expasy.org/scanprosite/) and Prosite profile PS50893; $(\ )_2$ indicates that the *T. urticae* ABC protein is a full transporter

[2] N-glycosylation sites were predicted using NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/); only N-glycosylation site with a "potential" score > 0. 05 and with a jury agreement were taken into account.

[3] O-glycosylation sites were predicted using NetOGlyc 3.1 server (Julenius *et al.* 2005) ; if the G-score was higher than 0.5 the residue was considered to be O-glycosylated; in this table the total number of O-glycosylated sites (glycosylated serines and threonines) is shown

Table V.2 - Subfamilies of ABC genes in *S. cerevisiae*, *C. elegans*, *H. sapiens*, *D. melanogaster*, *T. castaneum*, *D. pulex* and *T. urticae

ABC subfamily	*S. cerevisiae*	*C. elegans*	*H. sapiens*	*D. melanogaster*	*T. castaneum*	*D. pulex*	***T. urticae***
A	0	7	12	10	9	4	10
B-full	1	14	4	4	4	2	2
B-half	3	10	7	4	2	5	2
C	6	9	12	14	35	7	39
D	2	5	4	2	2	3	2
E	2	1	1	1	1	1	1
F	6	3	3	3	3	4	3
G	10	9	5	15	13	24	23
H	0	0	0	3	3	15	22
TOTAL	30	58	48	56	73	65	**103**

* numbers were derived from (Sturm *et al*. 2009; Liu *et al*. 2011b; Broehan *et al*. 2013) and this study; one additional ABCG gene was identified for *D. pulex* (Dappu1_300887, http://wfleabase.org/) in this study resulting in 65 *D. pulex* ABCs; a *C. elegans* ABCB half transporter discussed by (Gonzalez-Cabo *et al*. 2011) was added to the number of *C. elegans* ABCB half transporters resulting in 58 ABC transporters for *C. elegans*

Table V.3 - *T. urticae* ABC fragments

ABC-transporter fragments	Length (AA)	best BLASTp hit with complete ABC transporter	ABC subfamily	E-Value
tetur186g00040	112	tetur11g04030	B	1,00E-73
tetur186g00050	161	tetur11g04030	B	1,00E-54
tetur01g15320	129	tetur01g15330	C	2,00E-64
tetur02g15307	32	tetur09g00590	C	6,00E-06
tetur03g07560	310	tetur03g07490	C	7,00E-61
tetur03g09520	171	tetur03g09880	C	8,00E-50
tetur03g10123	78	tetur09g04620	C	2,00E-05
tetur03g10133	28	tetur03g09880	C	4,00E-09
tetur04g05510	196	tetur04g05540	C	8,00E-53
tetur04g05600	60	tetur04g05540	C	1,00E-19
tetur04g05620	234	tetur04g05540	C	1,00E-56
tetur04g09607	60	tetur04g05540	C	3,00E-16
tetur06g00220	693	tetur06g00360	C	0.0
tetur06g00280	941	tetur06g00360	C	0.0
tetur16g03390	119	tetur16g03480	C	2,00E-20
tetur17g00050	83	tetur14g02290	C	5,00E-11
tetur22g03023	68	tetur25g01780	C	6,00E-17
tetur24g02767	66	tetur14g02310	C	1,00E-05
tetur25g02132	36	tetur25g01780	C	4,00E-22
tetur28g02631	52	tetur01g07880	C	1,00E-17
tetur28g02641	168	tetur01g07880	C	2,00E-20
tetur31g00460	174	tetur25g01780	C	6,00E-24
tetur467g00010	177	tetur04g04360	C	1,00E-79
tetur55g00020	202	tetur01g07880	C	5,00E-31
tetur441g00040	45	tetur35g01360	D	3,00E-25
tetur11g02160	464	tetur32g00490	F	5,00E-175
tetur26g02853	33	tetur32g00490	F	4,00E-05
tetur02g15297	40	tetur04g06390	H	6,00E-11
tetur04g06420	158	tetur04g06390	H	5,00E-65
tetur04g09617	52	tetur18g00230	H	9,00E-16
tetur05g05040	101	tetur05g05000	H	3,00E-18
tetur05g09355	66	tetur05g05000	H	5,00E-36
tetur05g09365	92	tetur05g05000	H	5,00E-46
tetur19g01690	152	tetur19g01780	H	4,00E-72
tetur19g01710	297	tetur19g01780	H	1,00E-70
tetur19g01730	218	tetur19g01780	H	1,00E-151
tetur19g03421	56	tetur19g01780	H	3,00E-18
tetur19g03431	144	tetur19g01780	H	3,00E-64
tetur77g00040	195	tetur04g06390	H	1,00E-86
scaffold 30: 884878..884940 *	21	tetur30g01960	A	2,00E-06
scaffold 266: 382..438, 441..512*	19	tetur04g05540	C	2,00E-13
scaffold 211: 25-141*	39	tetur04g05540	C	1,00E-16

* no gene model could be constructed for these loci

3.1.1 The ABCA subfamily

The ABCA family comprises 9 full transporters (hereafter FTs) in *T. urticae.* In contrast to plants and some insects, no ABCA half transporters (hereafter HTs) were identified (Table V.1) (Verrier *et al*. 2008; Liu *et al*. 2011b; Xie *et al*. 2012). The *T. urticae* ABCA subfamily contains the largest *T. urticae* ABC protein, tetur25g01640 (2,302 amino acids) (Table V.1). ABCAs share a distinct set of characteristics across species: an extracellular loop between the first and second transmembrane helices (TMs), a conserved motif downstream of each NBD (Peelman *et al.* 2003) and a conserved motif at the N-terminus (xLxxKN, (Beers *et al.* 2011)). All *T. urticae* ABCAs have a characteristic extracellular loop between the first and the second TMs of each TMD (see Appendix V-E for position of TMs). The conserved motif downstream of each NBD was, except for tetur25g01640, present in all *T. urticae* ABCAs, while the N-terminus conserved motif could only be found in a single *T. urticae* ABCA (tetur25g01640). Instead of xLxxKN, the remainder of *T. urticae* ABCAs harbor either a xMxxKD/S (7) or xLxxHR (1) N-terminal motif. A phylogenetic analysis of metazoan ABCAs is shown in Fig. V.3. Six *T. urticae* ABCAs (tetur01g00580, tetur11g05030, tetur11g05040, tetur11g05200, tetur15g01990 and tetur30g01960) clustered together with high bootstrap support. These six ABCAs show high amino acid identity (62.3-94.1%, Appendix V-F) and have identical exonic structures (Appendix V-E), indicating they might have arisen by recent duplication events. Together with tetur01g15090, they form a sister-group with *D. melanogaster* CG31731, an ABCA reported to be down regulated in the salivary glands of an *E93* mutant of *D. melanogaster* ((Dutta 2008); the early ecdysone responsive gene *E93* is a critical regulator of programmed cell death during *D. melanogaster* metamorphosis). *D. melanogaster* CG31731 and the seven *T. urticae* ABCA genes cluster together, albeit with moderate bootstrap support, with a group of *C. elegans* ABCA transporters. The latter contains Ced-7, which is involved in the engulfment of cell corpses during programmed cell death in *C. elegans* (Wu and Horvitz 1998).

Further, tetur27g01890 and *D. melanogaster* CG34120 form a sister clade of human ABCA12 and ABCA13, while tetur25g01640, *D. pulex* Dappu1-312055 and Dappu1-312056 cluster with human ABCA1, ABCA2, ABCA4 and ABCA7 (Fig. V.3). These human ABCAs contain conserved predicted N-glycosylation sites at N400, N1453 and N1637 of human ABCA1 (Peelman *et al.* 2003). In addition, it has been experimentally shown that *D. melanogaster* CG34120 is also glycosylated at an asparagine (N272, at NASFEEL motif of CG34120 (Baycin-Hizal *et al.* 2011)) aligning with one of these conserved sites (N400

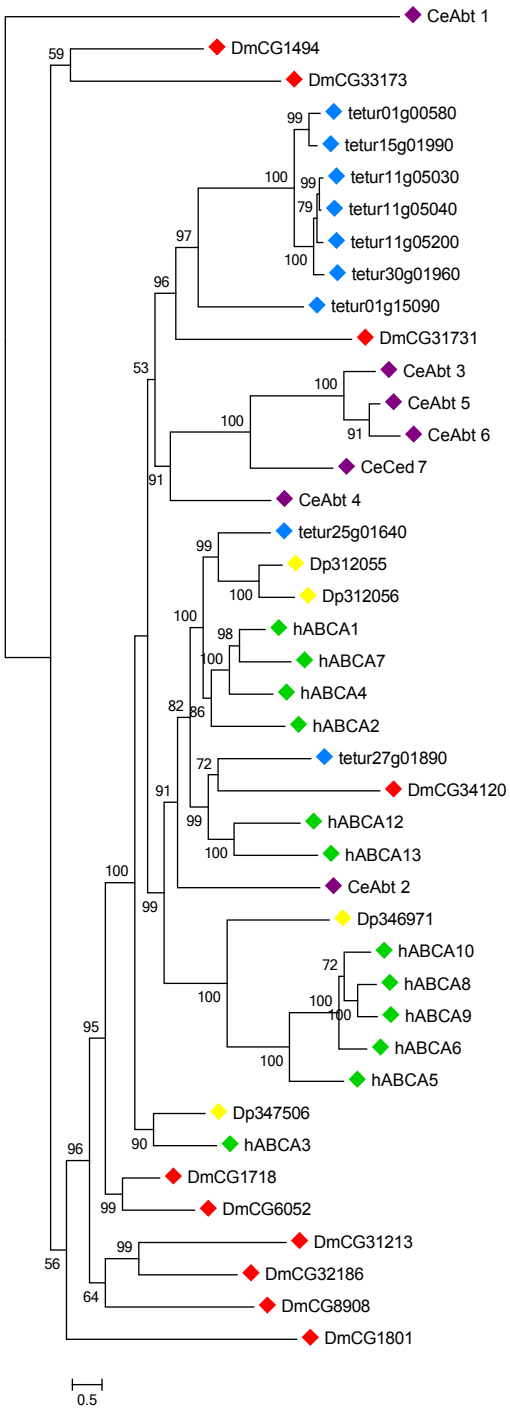


Figure V.3 - Phylogenetic analysis of ABCA proteins of five metazoan species.
Full-length ABC proteins were aligned using MUSCLE (Edgar 2004) and subjected to a maximum likelihood analysis using Treefinder (Jobb *et al.* 2004). The resulting tree was midpoint rooted. Numbers at the branch point of each node represent the bootstrap value resulting from 1000 pseudoreplicates (LR-ELW). Species, abbreviations, and color codes are: Ce, *C. elegans* (purple); h, *H. sapiens* (green); Dm, *D. melanogaster* (red); Dp, *D. pulex* (yellow); and tetur, *T. urticae* (blue). Accession numbers of metazoan ABC transporter sequences can be found in Appendix V-C while amino acid sequences of *T. urticae* ABC transporters can be found in Appendix V-D. The scale bar represents 0.5 amino-acid substitutions per site.

of human ABCA1). The tetur27g01890 and tetur25g01640 proteins also have many predicted N-glycosylation sites (Table V.1) of which at least one (N303 in tetur27g01890; N1705 in tetur25g01640) is shared with those conserved in human ABCA1, 2, 4, 7 and 12. In humans, these ABCAs have highly specialized roles in phospho- and sphingolipid export (Peelman *et al.* 2003). For example, human ABCA1 controls the initial steps leading to high-density lipoprotein (HDL) formation at the cell membrane and is crucial for reverse cholesterol transport from peripheral tissues to the liver (Zarubica *et al.* 2007). Human ABCA12 works as an epidermal keratinocyte lipid transporter and a defective ABCA12 results in loss of the skin lipid barrier (Akiyama *et al.* 2005; Albrecht and Viturro 2007). Although we cannot assign such highly specific roles to the two *T. urticae* ABCA orthologues above, they may also be involved in lipid transport processes.

3.1.2 The ABCB subfamily

The ABCB subfamily consists of 2 FTs and 2 HTs in *T. urticae* (Table V.1). A phylogenetic analysis of ABCB FTs revealed that transporters of each species in the analysis clustered into separate clades, confirming an earlier hypothesis by Sturm *et al*. (2009) that this subfamily has diversified through lineage-specific duplications (Fig. V.4*A*). This diversification hypothesis is supported in mites by the fact that the *T. urticae* ABCB FTs, tetur11g04030 and tetur11g04040, have well-supported phylogenetic clustering, similar exon patterns (15 and 16 exons for *tetur11g04030* and *tetur11g04040*, respectively (see Appendix V-E)) and high amino acid identity (64.8%, see Appendix V-F). *T. urticae* ABCB FTs form a sistergroup to a clade of *C. elegans*, *H. sapiens* and *D. melanogaster* ABCB FTs. The function of most members of this clade has been well documented in literature. Human ABCB FT, originally termed P-glycoproteins (P-gps) but now also known as multiple drug resistance (MDR) proteins, are among the best characterized ABC pumps and have been shown to be involved in transport of hydrophobic substrates including drugs, lipids, steroids, xenobiotics and peptides (Dean *et al.* 2001). The precise role of their orthologues in *Drosophila* has been a focus of recent study. *D. melanogaster* Mdr65 has been shown to function as an orthologue of human ABCB1/MDR1, a major ABC transporter of cytotoxic xenobiotics at the human blood-brain barrier, and is required for chemical protection of the fruitfly brain (Mayer *et al.* 2009) while Mdr49 has been shown to be essential in germ cell migration (Ricardo and Lehmann 2009). Interestingly, arthropod ABCB FT orthologues have frequently been linked to pesticide resistance (Buss and Callaghan 2008; Labbé *et al.* 2011). For example, inhibition of a *H. virescens* orthologue of human ABCB1 by the P-gp inhibitor quinidine decreased the

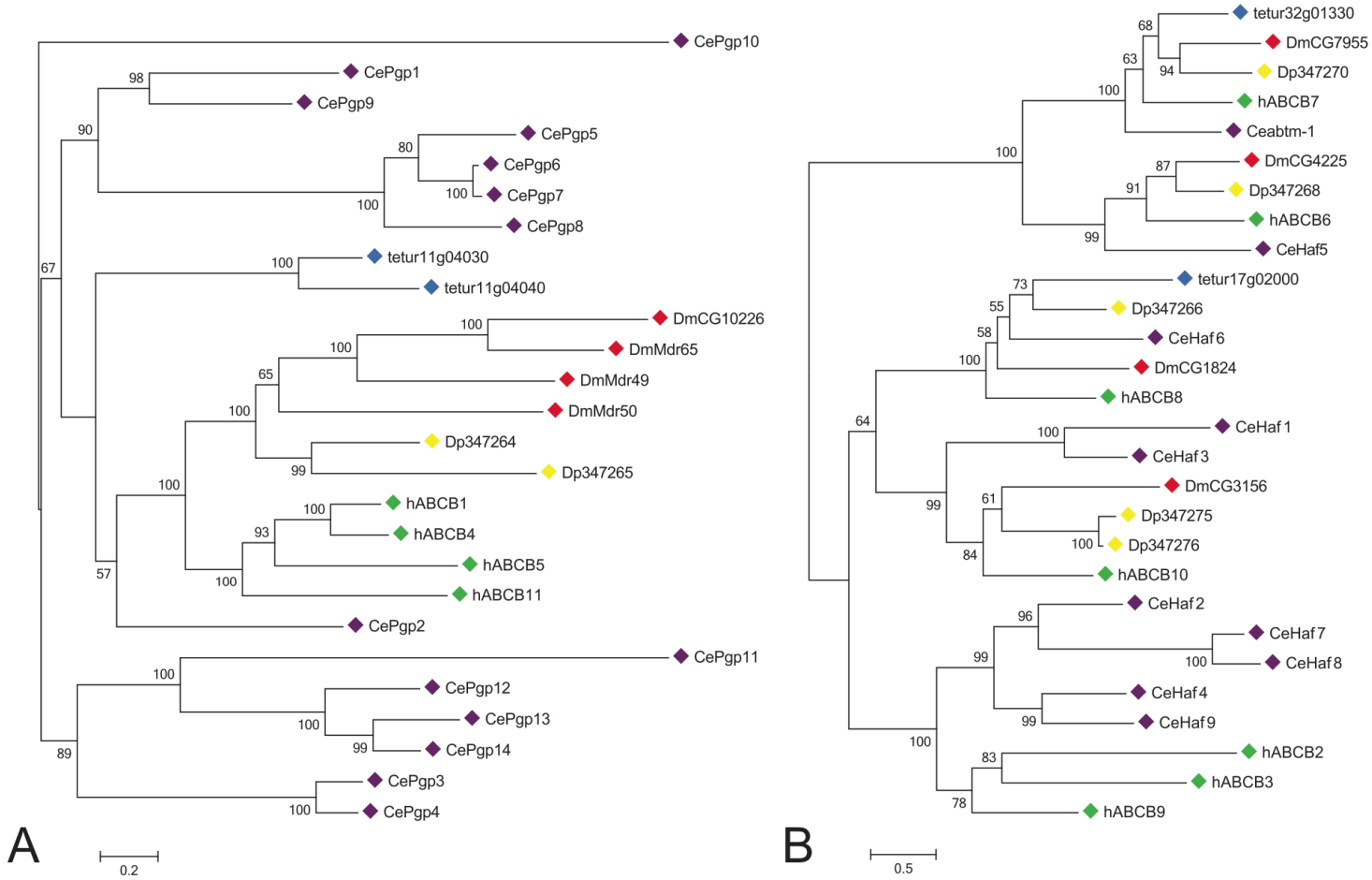


Figure V.4 - Phylogenetic analysis of ABCB proteins of five metazoan species

See the legend of Fig. V.3 for procedure, display details and abbreviations. (A) ABCB full transporters, (B) ABCB half transporters. The scale bar represents 0.2 and 0.5 amino-acid substitutions per site in Fig. V.4*A* and 4*B*, respectively.

toxicity of thiodicarb by 12.5-fold in a resistant strain, compared to 1.8-fold in a susceptible strain (Lanning *et al.* 1996). Recently, it was found that pretreatment of *D. melanogaster* with the P-gp inhibitor verapamil reduced the toxicity of DDT by 10-fold in a resistant strain (Strycharz 2010). The involvement of P-gps in pesticide resistance is probably best documented for ivermectin resistance. This compound has been shown to be a substrate for both mammalian as insect P-gps and several cases of P-gp associated ivermectin resistance have been reported (Buss and Callaghan 2008; Pohl *et al.* 2011; Yoon *et al.* 2011; Lespine *et al.* 2012).

A phylogenetic analysis of ABCB HTs revealed, as was also shown by Sturm *et al*. (2009), clear orthologous relationships between ABCB HTs, suggesting they have evolutionary conserved roles in metazoan species (Fig. V.4*B*). In the case of *T. urticae* ABCB HTs, an orthologous relationship between tetur32g01330 and *D. melanogaster* CG7955, *D. pulex* Dappu1-347270, *C. elegans* ABTM-1 and human ABCB7 was found, while tetur17g02000 groups together with *D. melanogaster* CG1824, *D. pulex* Dappu1-347266, *C. elegans* Haf-6, and human ABCB8. As both tetur32g01330 and tetur17g02000 are predicted (data not shown) to have a mitochondrial targeting signal, these *T. urticae* transporters are most likely trafficked to the mitochondria, as has been demonstrated for their human orthologues (human ABCB7 and ABCB8) (Zutz *et al.* 2009). This suggests that tetur32g01330 and tetur17g02000 fulfill a similar role as their human orthologues. The human ABCB7 protein plays a crucial role in iron homeostasis in the cytoplasm and mutations in this gene have been linked to several diseases (Zutz *et al.* 2009). Recently, it was also shown that disruption of the *C. elegans* orthologue (ABTM-1) of human ABCB7 induced oxidative stress and premature cell death (Gonzalez-Cabo *et al.* 2011). Furthermore, an orthologue (GenBank acc. No. AAEL006717) of human ABCB7 in the dengue vector, *Aedes aegypti*, was reported to be upregulated in an insecticide resistant strain (Liu *et al*. 2011b; Bariami *et al*. 2012). The function of human ABCB8 is not well understood, but it was shown to mediate resistance against the chemotherapeutic agent doxorubicin in melanoma cells (Elliott and Ai-Hajj 2009; Sachrajda and Ratajewski 2011). Also, Ichikawa *et al*. (2012) found that disruption of the mouse orthologue of human ABCB8 lead to cardiomyopathy and decreased mitochondrial iron export. Orthologues of the remaining human mitochondrial transporters, ABCB6 and ABCB10, could be identified in *D. pulex*, *D. melanogaster* and *C. elegans* but were not found in *T. urticae* (Dean *et al.* 2001; Sturm *et al.* 2009) (Fig. V.4*B*). Interestingly, the localization of human ABCB6 is currently under debate, as some studies suggest that it is located in

lysosomes (Kiss *et al*. 2012b). Finally, similarly to Sturm *et al*. (2009), we did not identify arthropod orthologues of human ABCB HTs related to antigen processing (human ABCB2, ABCB3 and ABCB9) (Fig. V.4*B*).

3.1.3 The ABCC subfamily

The ABCC subfamily consists of 39 transporters in *T. urticae*. To our knowledge, this is the largest number of ABCC transporters reported in any metazoan species, including the flour beetle, *T. castaneum,* which also has an exceptionally large number of ABCCs (Liu *et al*. 2011b; Broehan *et al*. 2013) (Table V.2). ABCC proteins are FTs, bearing 2 TMDs and 2 NBDs, with diverse functions: ion transport, cell-surface receptor activity and translocation of a broad array of substrates like drugs, cyclic nucleotides, endogenous compounds and their glutathione conjugates and glutathione (Kruh and Belinsky 2003; Jordan *et al.* 2008; Akrouh *et al.* 2009; Fukuda and Schuetz 2012). Because of their ability to extrude drugs, many of these ABCCs are also termed multidrug resistance associated proteins (MRPs) (Dean *et al.* 2001). MRPs can be categorized according to the presence or absence of a third N-terminal transmembrane-spanning domain (TMD_0). In humans, "long" MRPs like ABCC1, 2, 3, 6 and 10 (MRP1, 2, 3, 6 and 7, respectively) have such a TMD_0 while "short" human MRPs like ABCC4, 5, 11 and 12 do not (MRP4, 5, 8 and 9, respectively) (Deeley *et al.* 2006). In addition to MRPs, the ABCC family also harbors the cystic fibrosis transmembrane conductance regulator (CFTR, human ABCC7) and sulfonylurea receptors (human ABCC8/SUR1 and ABCC9/SUR2) (Dean *et al.* 2001).

In our phylogenetic analysis, 23 *T. urticae* ABCCs clustered with *D. melanogaster* CG6214, *D. pulex* Dappu1-347281 and a group of human "long" MRPs (MRP1, 2, 3 and 6) (Fig. V.5). Twenty-two of the transporters (size ranging from 1481 to 1529 amino acids) from this *T. urticae* ABCC clade also have a TMD_0, while we could not identify this additional transmembrane domain in tetur28g01950 (1287 amino acids) (Table V.1, Appendix V-E). Within this *T. urticae* ABCC clade, transporters clustered into 2 distinct groups. One group ("group 1") consists of 8 *T. urticae* ABCC genes, each with 7 introns, and *tetur04g04360*, which has only 4 introns (see Appendix V-E). Interestingly, tetur04g04360 is positioned at the basal node, indicating that intron gain events may have occurred in this group (Fig. V.5). The other group ("group 2") consists of 14 *T. urticae* ABCC genes: 12 with 10 introns, *tetur25g01780* with 11 introns and *tetur28g01950* having only 6 introns (Appendix V-E).

The intron loss in *tetur28g01950* can be directly linked to the lack of the TMD_0 (see above, Appendix V-E), while the nature of the intron gain event in *tetur25g01780* is unclear (the extra intron contains stop codons in each frame, and is located in a non conserved region). Human MRP1, MRP2, MRP3 and MRP6 along with *D. melanogaster* CG6214 and *D. pulex* Dappu1-347281 form a sister clade of *T. urticae* ABCC groups 1 and 2 (Fig. V.5). These proteins have been extensively studied as transporters of natural product drugs like anthracyclines and plant alkaloids (Borst *et al.* 2000). MRPs also share many substrates with human P-gps (see above) but, while P-gps transport drugs in their original form, MRPs mostly transport their glucuronate, sulfate and glutathione (GSH) conjugates. In the latter case, GSH is fused by GSTs with xenobiotics or their metabolites and finally transported out of the cell by MRPs (Ishikawa 1992; Borst *et al.* 2000; Xiong *et al.* 2010). In humans, GSTs from the alpha, mu and pi-class have been reported to act in synergy with MRPs (Sau *et al.* 2010). Intriguingly, a clear expansion of GSTs in the *T. urticae* genome was found for the delta and mu GST subclasses, the latter of which is not present in insects and until recently was believed to be vertebrate specific (Grbić *et al.* 2011; Reddy *et al.* 2011). Future studies should point out if there is a coordinated action between the GSTs of these subfamilies and the many MRP orthologues of *T. urticae*. Most of the biochemical properties of human MRP1 have been confirmed in its *D. melanogaster* orthologue CG6214 (Tarnay *et al.* 2004; Szeri *et al.* 2009). Recently, it was also found that exposure to the P450 mono-oxygenase inhibitor piperonylbutoxide and the antimetabolite/antifolate drug methotrexate alters the expression (27-fold and 1100-fold upregulation, respectively) of the *D. melanogaster* CG6214 gene in the Malpighian tubuli, which are organs known to play important roles in excretion and xenobiotic detoxification (Chahine and O'Donnell 2009; Chahine and O'Donnell 2011). The orthologue (*PhABCC4*) of CG6214 in the human body louse, *Pediculus humanus*, has also been reported to be upregulated after exposure to the pesticide ivermectin. Moreover, injection of *PhABCC4* dsRNA in *P. humanus* female lice increased their sensitivity to ivermectin by 20-30% (Yoon *et al.* 2011).

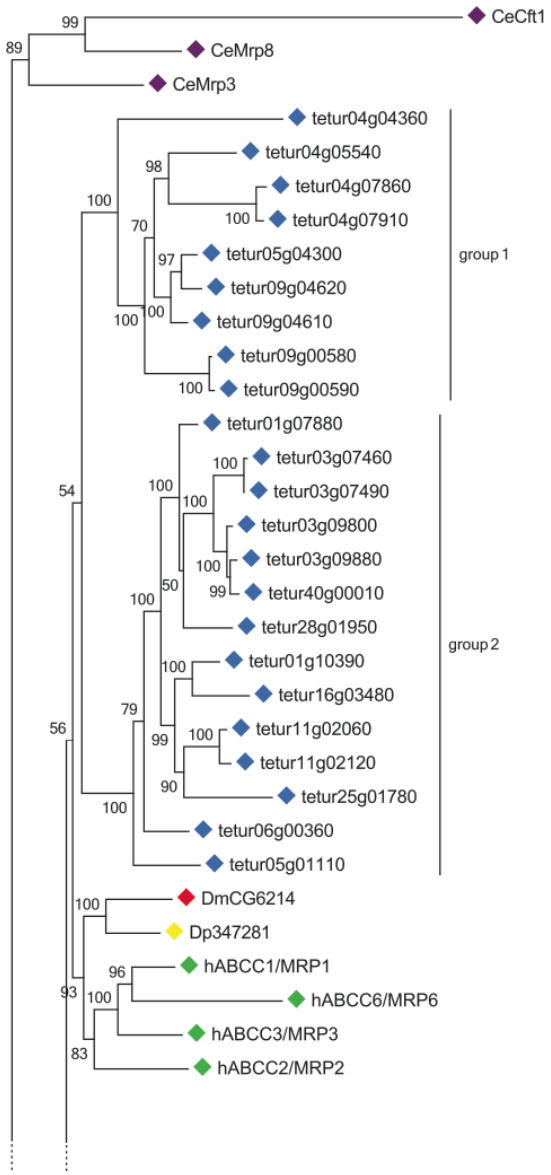

CeCft1
CeMrp8
CeMrp3
tetur04g04360
tetur04g05540
tetur04g07860
tetur04g07910
tetur05g04300
tetur09g04620
tetur09g04610
tetur09g00580
tetur09g00590
group 1
tetur01g07880
tetur03g07460
tetur03g07490
tetur03g09800
tetur03g09880
tetur40g00010
tetur28g01950
tetur01g10390
tetur16g03480
tetur11g02060
tetur11g02120
tetur25g01780
tetur06g00360
tetur05g01110
group 2
DmCG6214
Dp347281
hABCC1/MRP1
hABCC6/MRP6
hABCC3/MRP3
hABCC2/MRP2

Figure V.5

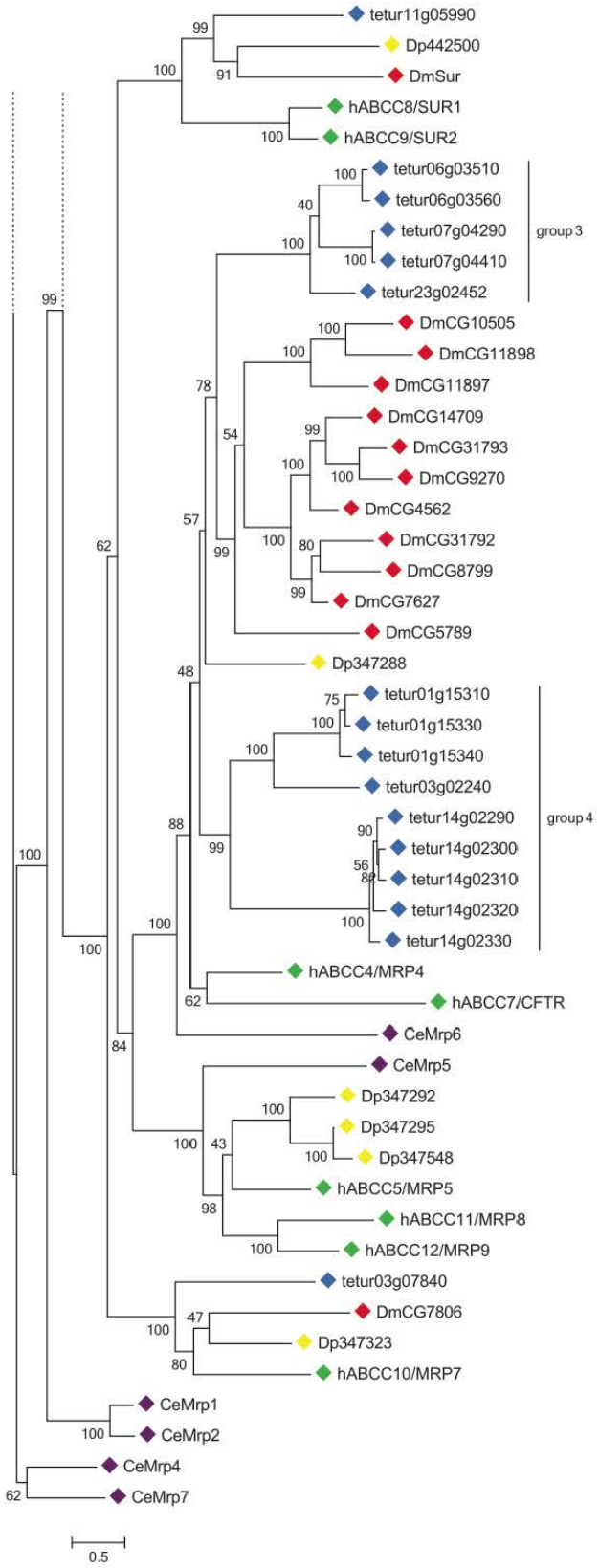


Figure V.5 (continued) - Phylogenetic analysis of ABCC proteins of five metazoan species.

See the legend of Fig. V.3 for procedure, display details and abbreviations.

Fourteen spider mite ABCCs clustered with high bootstrap support in a group with 11 *D. melanogaster* ABCC proteins, human ABCC4/MRP4 and ABCC7 and *D. pulex* Dappu1-347288 (Fig. V.5). All fourteen *T. urticae* ABCCs show the structural properties of "short" MRPs (see above, Table V.1, Appendix V-E). Five of these *T. urticae* ABCCs group as a sister clade ("group 3") of the 11 *D. melanogaster* ABCC proteins (Fig. V.5). Detailed physiological roles for most of the 11 *D. melanogaster* ABCC proteins are unknown. *D. melanogaster* CG10505 is regulated by heavy metals through the metal-responsive transcription factor 1 (MTF-1) and contributes to metal homeostasis (Yepiskoposyan *et al.* 2006). *D. melanogaster* CG14709 controls responsiveness to O_2 deprivation and might also be involved in oxidative stress response (Monnier *et al.* 2002; Huang and Haddad 2007) while 31% of embryos of *D. melanogaster* CG7627 mutants were unable to heal wounds 16h postwounding (Campos *et al.* 2010). Recently, it was also shown that *D. melanogaster* CG4562, which is highly expressed in the midgut, was upregulated in *Cyp6g1* knockdown flies, indicating molecular crosstalk within a detoxification network (Shah *et al.* 2012). Furthermore, a point mutation in lepidopteran homologs of this *D. melanogaster* ABCC group has been clearly linked with resistance against the *B. thuringiensis* Cry1A toxin (Baxter *et al.* 2011; Heckel 2012). Another 9 *T. urticae* ABCCs ("group 4") form, together with *D. pulex* Dappu1-347288 and the 11 *D. melanogaster* and five *T. urticae* ABCCs ("group 3", see above) a sister clade of human ABCC4 and ABCC7 (Fig. V.5). The function of *D. pulex* Dappu1-347288 is not known, but human ABCC7/CFTR acts as a chloride channel, a unique function not found in any other ABC transporter (Jordan *et al.* 2008). In our analysis, CFTR clustered with human ABCC4/MRP4, which is its closest ABC paralog according to Jordan *et al*. (2008). Human MRP4 (and MRP5) have the ability to transport a range of endogenous molecules involved in cellular signaling, like cyclic nucleotides, eicosanoids and conjugated steroid hormones (Fukuda and Schuetz 2012). As a drug transporter, MRP4 also stands out for its broad substrate specificity, covering antiviral, antibiotic, cardiovascular and cytotoxic agents (Russel *et al.* 2008).

Interestingly, many transporter genes in *T. urticae* ABCC groups 1-4 form closely related sister groups and show high amino acid identity between their corresponding protein sequences (see Appendix V-F). Together with their conserved exon pattern (see Appendix V-E), this strongly suggests that multiple tandem duplications underlie the proliferation of these genes. This is in contrast to the crustacean *D. pulex*, which has only few ABCC genes, (Sturm *et al.* 2009) and more closely resembles what is generally observed for insects, where gene

duplication of ABCC genes occurs frequently (Liu *et al*. 2011b; Xie *et al*. 2012; Broehan *et al*. 2013).

Furthermore, clear orthologous relationships were found for the two remaining *T. urticae* ABCCs: tetur03g07840 and tetur11g05990. Tetur03g07840 clustered as an orthologue of *D. melanogaster* CG7806, *D. pulex* Dappu1-347323 and human ABCC10/MRP7 (Fig. V.5). Similar to its orthologues, tetur03g07840 has a TMD_0 (Hopper-Borge *et al.* 2009; Sturm *et al.* 2009) (Table V.1, Appendix V-E). The functions of *D. melanogaster* CG7806 and *D. pulex* Dappu1-34723 are not known. A growing understanding of the physiological role of human ABCC10/MRP7 is, on the other hand, beginning to emerge. Human MRP7 is distinct from other human ABCCs in that it shows little or no activity towards glutathione, sulfate conjugates and cyclic nucleotides, substrates that can be handled by other human MRPs (see above). Instead, human MRP7 is able to confer resistance to taxanes (which are diterpenes originally derived from plants of the genus *Taxus* and widely used as chemotherapy agents) (e.g. docetaxel) (Hopper-Borge *et al.* 2009). However, the presence of one-to-one orthologues in other metazoans, might indicate a more conserved function of this ABCC protein in this group of species.

Tetur11g05990 is located in the same clade as *D. pulex*, human and *D. melanogaster* sulfonylurea receptors (SURs). In contrast to vertebrates, the N-terminal SUR Interpro-motif (IPR000388) is not present in tetur11g05990 and other arthropod SURs (see OrthoDb group EOG531ZCW, (Waterhouse *et al.* 2011)). However, the presence of a TMD_0 typical for SURs and "long" MRPs (Deeley *et al.* 2006), the relatively high amino acid similarity (55.1 and 54.0%, respectively) and the well-supported clustering with human ABCC8/SUR1 and ABCC9/SUR2 all support the idea that tetur11g05990 is a SUR homologue. Four SUR subunits assemble into an octameric complex with four pore-forming subunits, characteristic for inwardly rectifying potassium (K_{ir}) channels, to form ATP-sensitive potassium (K_{ATP}) channels (Akrouh *et al.* 2009). Three orthologues of these pore- forming subunits were also found in the *T. urticae* genome (tetur17g01380, tetur24g01270 and tetur24g01280 having a BLASTx E-value of $1e^{-99}$, $4e^{-103}$ and $3e^{-103}$ with *Ir* (*CG44159*) of *D. melanogaster*), suggesting that a functional K_{ATP} channel can be formed in *T. urticae*. K_{ATP} channels are involved in multiple physiological processes, with roles in glucose homeostasis, ischemic protection and innate immunity (Akrouh *et al.* 2009; Eleftherianos *et al.* 2011). Intriguingly, in 2004 it was suggested that the SUR was the direct target of benzoylureas, a group of chitin

synthesis inhibitors (Abo-Elghar *et al.* 2004). This was largely based on similar effects of glibenclamide, a well-known SUR inhibitor in humans and anti-diabetic drug, on the inhibition of chitin synthesis. However, it was later shown by Gangishetti *et al*. (2009) that SUR is not expressed in the *D. melanogaster* epidermis, where chitin disruption is observed. Recently, based on genetic mapping of etoxazole resistance genes, it was suggested that the action of chitin synthesis inhibitors is mediated by a direct interaction with chitin synthase, a processive glycosyl transferase (Van Leeuwen *et al.* 2012). The lack of a role for SUR in chitin production, transport or metabolism is further confirmed by recent studies, where it was shown that the SUR is dispensable for chitin synthesis in *D. melanogaster* (Meyer *et al.* 2012), and RNAi knockdown of its orthologue in *T. castaneum* did not result into a phenotype (Broehan *et al.* 2013). Elucidating the role of SUR in *T. urticae* will therefore require additional studies.

Finally, no orthologues of human ABCC5, 11 and 12 were identified in *T. urticae*, although three orthologues were found in the genome of *D. pulex* (Fig. V.5), confirming earlier findings by Sturm *et al.* (2009). Surprisingly, a single nucleotide polymorphism in human ABCC11 was identified as the determinant of the human earwax type (Yoshiura *et al.* 2006). However, the potential roles of related transporters in other organisms (such as *D. pulex*) are not clear.

3.1.4 The ABCD subfamily

The ABCD subfamily harbors HTs that in humans are located in the peroxisome where they are involved in the import of long and branched chain acyl-coA into this organelle (Morita and Imanaka 2012). The *T. urticae* genome has 2 ABCD genes, *tetur05g06640* and *tetur35g01360* (Table V.1). This number of ABCD genes equals those found in insects (Liu *et al.* 2011b) while 3, 4 and 5 are found in the genomes of *D. pulex*, *H. sapiens* and *C. elegans*, respectively (Table V.1). *T. urticae* ABCDs carry the EAA-like motif between TM4 and TM5 and the loop1 motif (L105, R108 and T109 (*S. cerevisiae* numbering)), both considered to be essential for canonical ABCD function (Shani *et al.* 1996). The clear orthologous relationships we identified between *T. urticae* ABCDs and other metazoan ABCDs (Fig. V.6*A*) suggests that the function of *T. urticae* ABCDs is likely to be conserved with those in other metazoans.

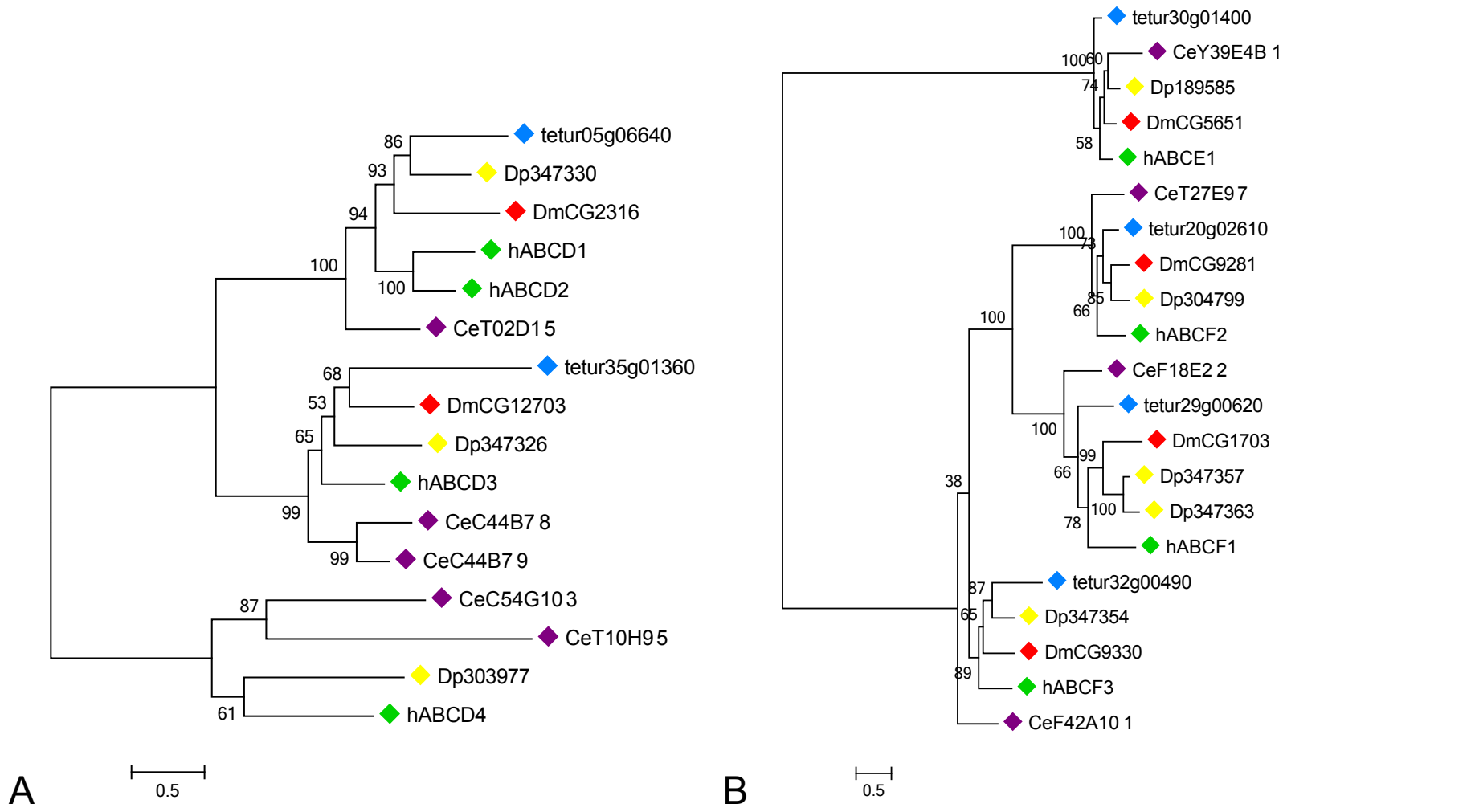


Figure V.6 - Phylogenetic analysis of ABCD, ABCE and ABCF proteins of five metazoan species.
See the legend of Fig. V.3 for procedure, display details and abbreviations. (A) ABCD subfamily (B) ABCE and –F subfamily

3.1.5 The ABCE and –F subfamily

The ABCE and F proteins are characterized by two linked NBDs, but lack TMDs and thus are involved in biological processes other than transport. The ABCE protein is essential in all eukaryotes examined to date and is one of the most conserved proteins known (Barthelme *et al.* 2011). Human ABCE1 was first discovered as an inhibitor of RNase L (Zhou *et al.* 1993), but was later found to have a more fundamental role in ribosome biogenesis and translation regulation (Barthelme *et al.* 2011). In line with all eukaryotes to date, we found one ABCE protein (tetur30g01400) in *T. urticae* (Table V.1 and V.2) that has high amino acid identity (77.1%) with *D. melanogaster* ABCE1 (pixie) (Appendix V-F). Similar to human ABCE, ABCF1 is involved in translation regulation, but probably does not play a role in ribosome biogenesis (Paytubi *et al.* 2009). In most eukaryotes 3 ABCF genes are found, and *T. urticae* conforms to this expectation (Liu *et al*. 2011b, Table V.1 and V.2). The essential role of ABCE and ABCF genes was recently shown in the flour beetle, *T. castaneum*, where RNAi–mediated knockdown of members of the ABCE and F families resulted in 100% mortality in penultimate larvae (Broehan *et al.* 2013).

A phylogenetic analysis of ABCE and ABCF proteins was performed together (Fig. V.6). Tetur30g01400 grouped with metazoan ABCE1 orthologues, while each *T. urticae* ABCF (tetur20g02610, tetur29g00620 and tetur32g00940) clustered into well-supported separate clades with its metazoan orthologues, *C. elegans* F42A10.1 excluded (Figure V.6*B*). The ABCE and ABCF subfamilies are highly conserved, and *T. urticae* ABCE1 and ABCFs probably have analogous roles as their orthologues in other metazoans.

3.1.6 The ABCG subfamily

The ABCG transporter family is present in most metazoan species, fungi and plants such as *Arabidopsis*. For metazoan species, only ABCG HTs have been reported to date, while in plants and fungi also ABCG FTs are present (Dean *et al.* 2001; Verrier *et al.* 2008; Kovalchuk and Driessen 2010). In humans, ABCG HTs are primarily implicated in transport of endogenous and dietary lipids, while the human ABCG2 functions as a multidrug efflux pump (Tarr *et al.* 2009; Kerr *et al.* 2011). Within the *T. urticae* genome we identified 23 ABCGs, all having a typical reverse domain organization (the NBD is localized to the N-terminal side of the TMD) (Table V.1, see Appendix V-E). A similar number of ABCGs has also been found in *D. pulex,* and is the highest reported among metazoan species (Sturm *et al.*

2009) (Table V.2). According to Sturm *et al*. (2009), the high number of ABCG genes in *D. pulex* and *D. melanogaster* genomes is due to extensive lineage specific duplications. Our phylogenetic analysis confirms this hypothesis not only for these two arthropod species but also for *T. urticae*, with twenty out of 23 ABCGs grouping into one of the two *T. urticae* specific clades (Fig. V.7).

One clade (bottom of tree, Fig. V.7) consists of 11 *T. urticae* ABCGs, each having a maximum of 1 intron (see Appendix V-E). Another *T. urticae* specific ABCG clade comprises 9 transporters of which 7 are located next to each other on scaffold 9. These seven ABCGs show high amino acid identity (between 50.8 and 91.7 %) and have a conserved exon pattern (9 exons), indicating a common origin by successive tandem duplication events. Together with tetur06g05430 and tetur02g11270 they form a well-supported sister clade of *D. melanogaster* white and its *D. pulex* orthologues. Interestingly, no orthologues of *D. melanogaster* ABCGs brown and scarlet were found in *T. urticae*, while only one *D. pulex* orthologue of scarlet could be identified (Fig. V.7, (Sturm *et al.* 2009)). About a century ago, the discovery of *D. melanogaster white* mutants with a remarkable eye-color phenotype marked the beginning of *Drosophila* genetics. As a consequence, *D. melanogaster white* is one of the most intensively studied fruit fly genes (Green 1996). *D. melanogaster* white dimerises with either *D. melanogaster* scarlet or brown to form a transporter involved in the uptake of pigment precursors (guanine and tryptophan) in cells of developing compound and simple eyes (Mackenzie *et al.* 1999). *T. urticae* has, in contrast to *D. pulex* and *D. melanogaster*, no compound eyes and only four simple eyes (ocelli) (Mills 1974). Although no *T. urticae* orthologues of scarlet or brown were identified, dimerisation between the nine *T. urticae* co-orthologues of *D. melanogaster* white might result in a transporter capable of translocating pigment precursors into the cells of the spider mite ocelli (which have red pigment, as in *D. melanogaster*). However, these transporters might also have other functions besides transporting pigment precursors, as in other species roles have been documented in courtship behavior (Zhang and Odenwald 1995; Anaka *et al.* 2008), transport of biogenic amines (Borycz *et al.* 2008) and uptake of uric acid (Tatematsu *et al.* 2011) as was shown for *D. melanogaster* white and/or its *B. mori* orthologue.

In the middle of the ABCG phylogenetic tree, tetur01g16280 clustered with human ABCG8, *D. melanogaster* CG31121 and *D. pulex* Dappu1-258299, while tetur01g16290 clustered with human ABCG5, *D. melanogaster* CG11069, and *D. pulex* Dappu1-300887. *C. elegans* orthologues of human ABCG5/8 could not be identified (Fig. V.7).

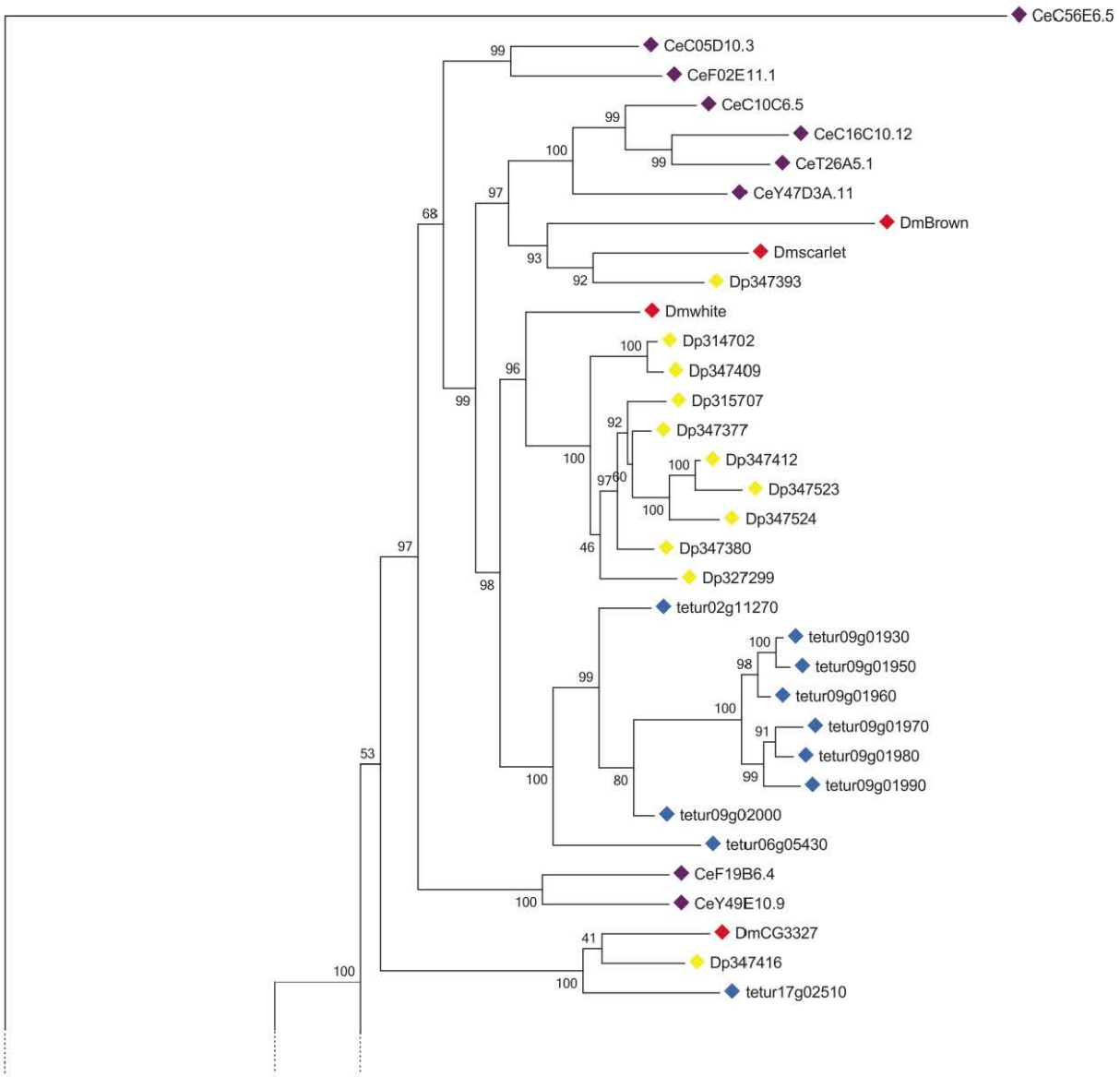

CeC56E6.5
CeC05D10.3
CeF02E11.1
CeC10C6.5
CeC16C10.12
CeT26A5.1
CeY47D3A.11
DmBrown
Dmscarlet
Dp347393
Dmwhite
Dp314702
Dp347409
Dp315707
Dp347377
Dp347412
Dp347523
Dp347524
Dp347380
Dp327299
tetur02g11270
tetur09g01930
tetur09g01950
tetur09g01960
tetur09g01970
tetur09g01980
tetur09g01990
tetur09g02000
tetur06g05430
CeF19B6.4
CeY49E10.9
DmCG3327
Dp347416
tetur17g02510

Figure V.7

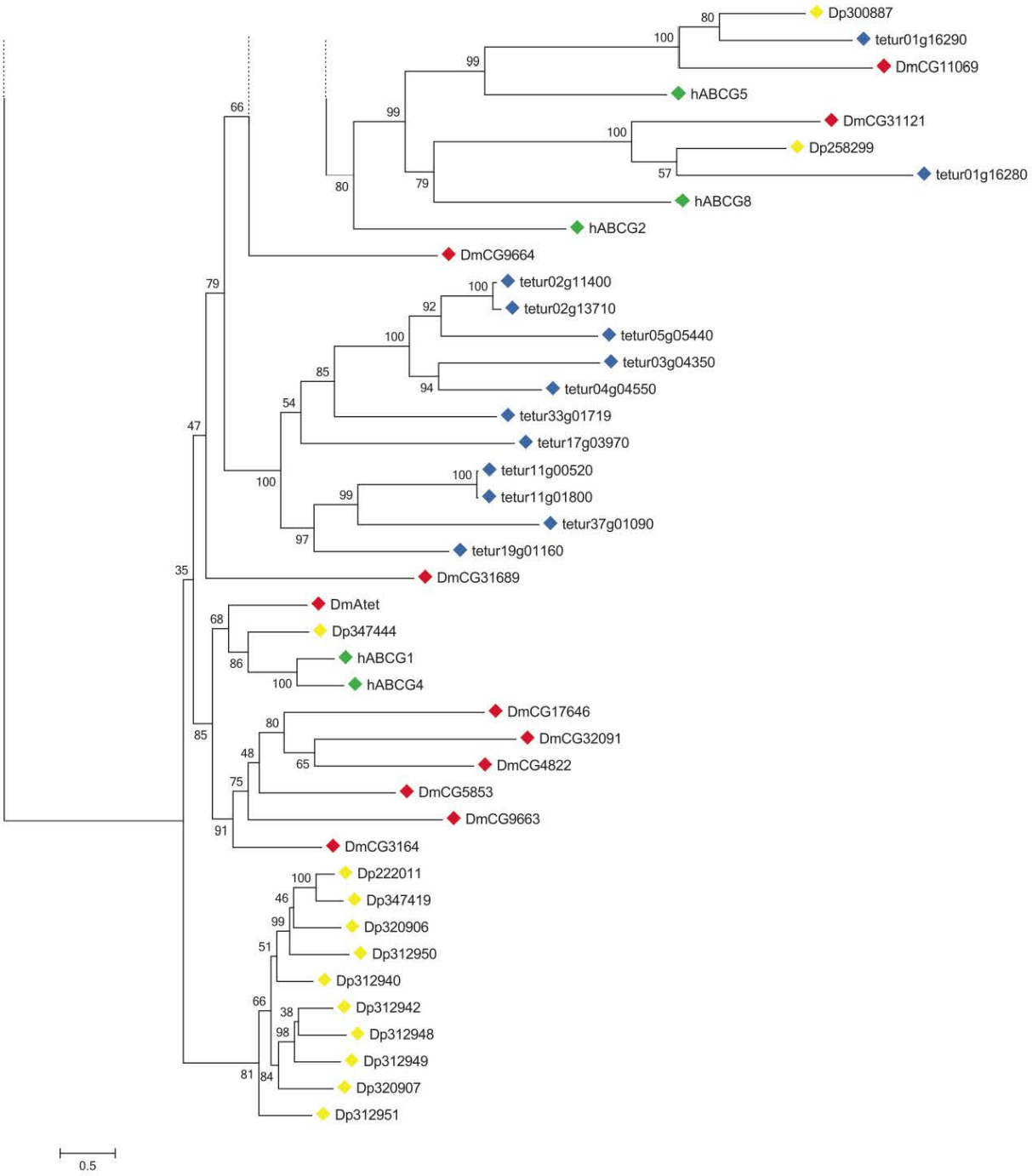


Figure V.7 (continued) - Phylogenetic analysis of ABCG proteins of five metazoan species.

See the legend of Fig. V.3 for procedure, display details and abbreviations.

Similar to human *ABCG5/8*, *D. melanogaster CG31121/CG11069* (McQuilton *et al.* 2012) and *D. pulex Dappu1-258299/Dappu1-300887* (http://wfleabase.org/), *tetur01g16280* and *tetur01g16290* are found juxtaposed in a head to head orientation. Annilo *et al*. (2006) suggested an evolutionary constraint on the separation of these genes, probably for the maintenance of shared regulatory regions. In humans, ABCG5 and ABCG8 are both glycoproteins and obligate heterodimers that limit intestinal absorption and promote biliary excretion of neutral sterols (Graf *et al.* 2003). Both tetur01g16280 and tetur01g16290 have at least one well-predicted glycosylation site (Table V.1). Together with their head-to-head arrangement and the well-supported clustering with human ABCG8 and 5, it seems likely that these *T. urticae* ABCGs have similar functions as their human counterparts.

A clear orthologous relationship was found between tetur17g02510, *D. melanogaster* CG3327 and *D. pulex* Dappu1-347416. *D. melanogaster* CG3327, also known as E23 (Early gene at 23), is a 20-OH ecdysone (20E) induced ABC transporter that is capable of regulating 20E responses during metamorphosis, probably by removing 20E from cells (Hock *et al.* 2000). Recently, Broehan *et al*. (2013) showed through RNAi-mediated knockdown experiments and expression profiling that the *T. castaneum* orthologue of E23 (TcABCG-8A) appears to serve a similar function in metamorphosis. In addition, it is also believed that E23 controls the circadian clock in adult flies through ecdysone-mediated expression of the clock gene *vrille* (Itoh and Matsumoto 2012). Interestingly, it was shown that not only the *B. mori* orthologue of *E23* (BmABC010557) but also four other midgut-specific *B. mori* ABCG genes (*BmABC005226*, *BmABC005203*, *BmABC005202* and *BmABC010555*) could be induced by 20E (Liu *et al*. 2011b). As *T. urticae* uses a different molting hormone (ponasterone A instead of 20E) compared to arthropods (Grbić *et al.* 2011), future experiments are required to establish if tetur17g02510 has a similar function as its insect counterparts, and more specifically whether it can be induced by ponasterone A.

T. urticae orthologues of human ABCG1 and 4 were not identified in the phylogenetic analysis of ABCG transporters, and only one was found in *D. melanogaster* (Atet) and *D. pulex* (Dappu1-34744) (Fig. V.7). The function of human ABCG4 is not well understood while it is proposed that human ABCG1 functions in conjunction with human ABCA1 and is involved in cholesterol homeostasis (Tarr *et al.* 2009; Kerr *et al.* 2011). The *D. melanogaster* orthologue of human ABCG1 (Atet) has been poorly characterized, but is expressed in the trachea (Kuwana *et al.* 1996). Finally, no clear orthologues of human ABCG2 were identified

(Fig. V.7). This transporter is the most thoroughly characterized human ABCG and is capable of transporting an array of substrates, including anticancer drugs (Tarr *et al.* 2009; Kerr *et al.* 2011; Fukuda and Schuetz 2012). Because of this feature, it has been proposed that arthropod ABCGs could be involved in pesticide resistance (Labbé *et al.* 2011). However, to the best of our knowledge there have only been two studies that correlated increased arthropod ABCG expression levels with resistance (Jones *et al.* 2012; You *et al.* 2013). In both reports, however, no functional evidence was presented. In Fungi on the other hand, several cases of ABCG FTs involved in fungicide resistance, have been reported (Lamping *et al.* 2010).

3.1.7 The ABCH subfamily

The ABCH subfamily was first discovered in *D. melanogaster* and is lacking in mammals, plants or fungi (Dean *et al.* 2001; Verrier *et al.* 2008; Kovalchuk and Driessen 2010). In addition to arthropods ((Dean *et al*. 2001; Sturm *et al*. 2009; Liu *et al*. 2011b; Broehan *et al*. 2013), this study), members of this subfamily have also been reported in teleost fish (Dean and Annilo 2005; Annilo *et al.* 2006; Popovic *et al.* 2010). Most insects have only 3 ABCH genes, while 15 and 22 are present in *D. pulex* and *T. urticae,* respectively ((Liu *et al*. 2011b; You *et al*. 2013), Table V.2). Liu *et al*. (2011b) suggested that all insect ABCHs diversified from a common ancestral copy. According to our phylogenetic analysis, this insect ancestor seems not to be shared with *T. urticae* ABCH proteins (Fig. V.8). *Tetranychus* ABCHs clustered, similar to *D. pulex* ABCHs (see also (Sturm *et al.* 2009)), into a distinct clade, indicating that the diversity of the ABCH family in *T. urticae* has been due to lineage specific duplications. Interestingly, 16 of the 22 *T. urticae* ABCHs appear to be intronless (Appendix V-E). Although ABCHs have the same structural organization as metazoan ABCGs (HTs, NBD at N-terminal side of TMD), their physiological functions have remained enigmatic. In the zebrafish, *Danio rerio*, ABCH1 has highest expression in brain, gills and kidney followed by lower expression in intestine, gonads, skeletal muscle and liver (Popovic *et al.* 2010). *D. melanogaster* ABCHs are enriched in the adult crop and hindgut (Robinson *et al.* 2013) and at least one of them (CG9990) is glycosylated as shown by mass spectrometry of *N*-glycosylated peptides (Baycin-Hizal *et al.* 2011). An RNAi screen of *D. melanogaster* genes revealed that an RNAi line that silences *CG9990* is lethal (Mummery-Widmer *et al.* 2009; Zhang *et al.* 2009). In addition, microarray analysis demonstrated an almost two-fold upregulation of a *D. melanogaster* ABCH (*CG33970*) after cold hardening of adult fruit flies (Qin *et al.* 2005). In the diamondback moth *Plutella xylostella*, it was recently found that an ABCH transporter (Px014955, http://iae.fafu.edu.cn/DBM) was the most up-regulated ABC

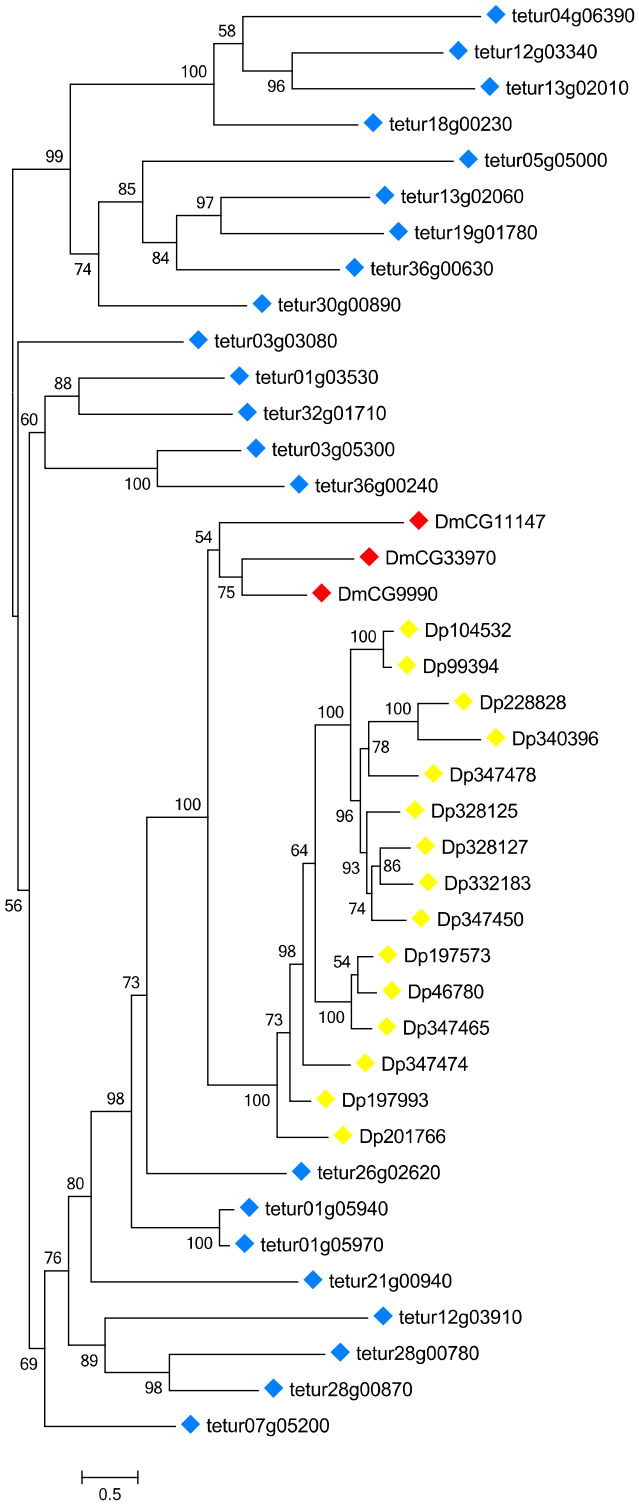


Figure V.8 - Phylogenetic analysis of ABCH proteins of five metazoan species.
See the legend of Fig. V.3 for procedure, display details and abbreviations.

gene in two resistant strains (You *et al.* 2013). The most groundbreaking finding about insect ABCH function was just recently reported by the excellent study of Broehan *et al*. (2013). RNAi-mediated knockdown of an ABCH gene (*TcABCH-9C*) in *T. castaneum* larvae resulted in dessication and 100% mortality. Injection of *TcABCH-9C* dsRNA into adults also drastically reduced the number of eggs laid and all eggs failed to hatch. Furthermore, cryosections of *TcABCH-9C* dsRNA injected larvae stained with Nile Red (a fluorescent dye that stains lipids) revealed a lack of lipids in the epicuticle. Based on these results, the authors suggested that TcABCH-9C functions as a transporter of lipids to the cuticle and is required for the formation of a waterproof barrier in the epicuticle.

3.2 Expression profile of ABC genes

We assessed expression of ABC genes across development in the *T. urticae* London reference strain, as well as in London after transfer from a benign host (bean, *Phaseolus vulgaris*) to two more challenging hosts (*Arabidopsis thaliana* and tomato, *Solanum lycopersicum*; (Grbić *et al.* 2011)). For the developmental and host transfer experiments, we used existing RNA-seq reads from, but we recalculated gene expression using newly described or corrected ABC gene models curated as part of this study. We further examined previously published microarray data to assess the expression profiles of ABC genes in two spider mite strains, MR-VP and MAR-AB, that are resistant to multiple pesticides (see Chapter IV).

As assessed by RNA-seq expression quantification, the majority of ABC genes were found to be expressed – 88 of the 103 full length *T. urticae* ABC genes had an RPKM of >1 in at least one of the spider mite life stages or on one of the plant hosts (Fig. V.9). In contrast, nearly all *T. urticae* ABC fragments or pseudogenes were not expressed (Fig. V.10). Most full-length *T. urticae* ABC genes for which we detected no expression across development or on different hosts belonged to either ABCA, C or G subfamilies, many of which are tandem duplicated genes (Table V.1, Fig. V.9). In *C. elegans*, tandem duplicated genes were shown to be subfunctionalized, with strong stage or tissue dependent expression (Zhao *et al.* 2004). Whether *T. urticae* ABC genes that lacked expression support in the existing data are expressed at low levels, in highly restricted expression domains, or alternatively are expressed under specific environmental conditions (i.e., host plants not included in this analysis), remains to be determined.

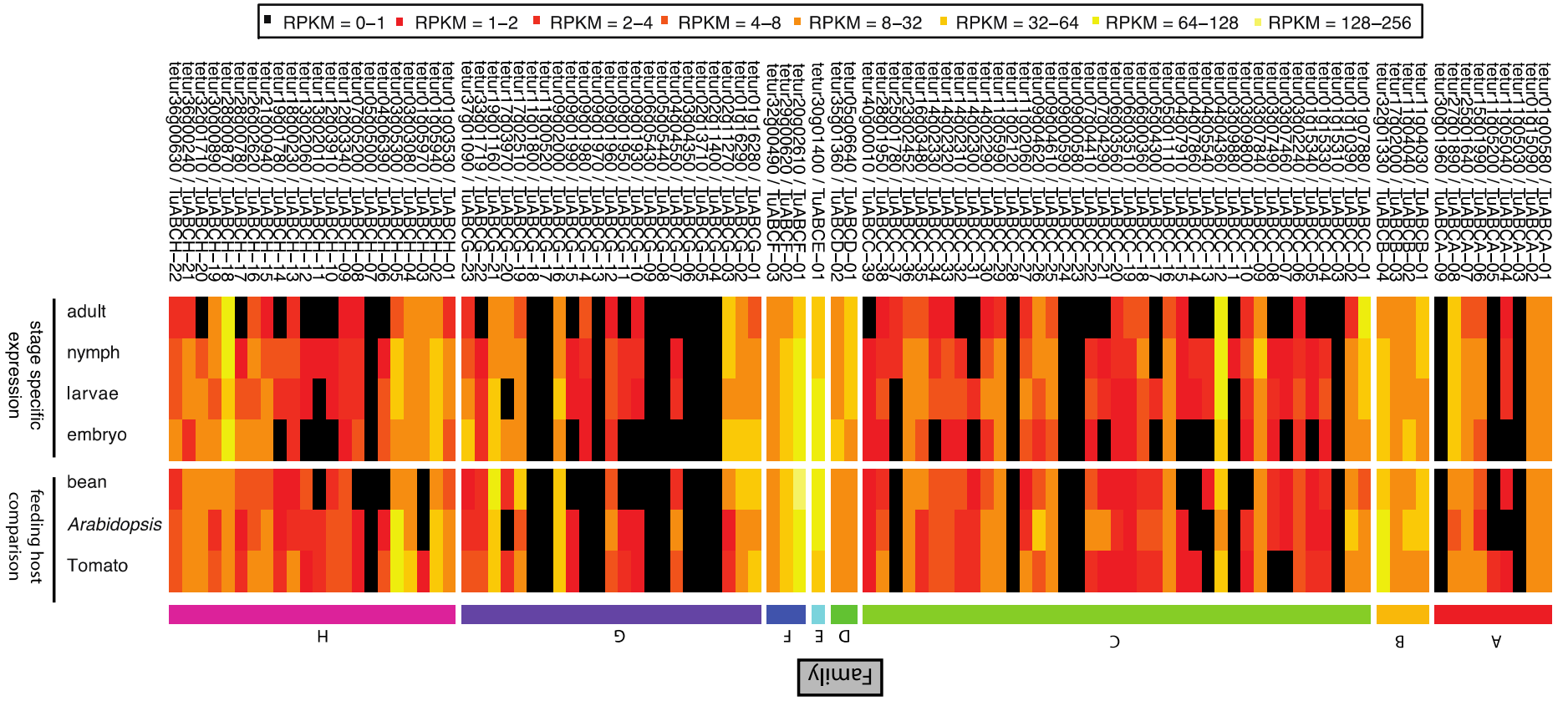


Figure V.9 – Heat map of expression values (RPKM values) of 103 *T. urticae* ABC genes of mites on different host plants (bean, tomato and *Arabidopsis*) and from four different life stages (embryo, larvae, nymph and adult).

ABC genes were considered as being expressed when they had an RPKM of >1 in at least one of the spider mite life stages or on one of the plant hosts.

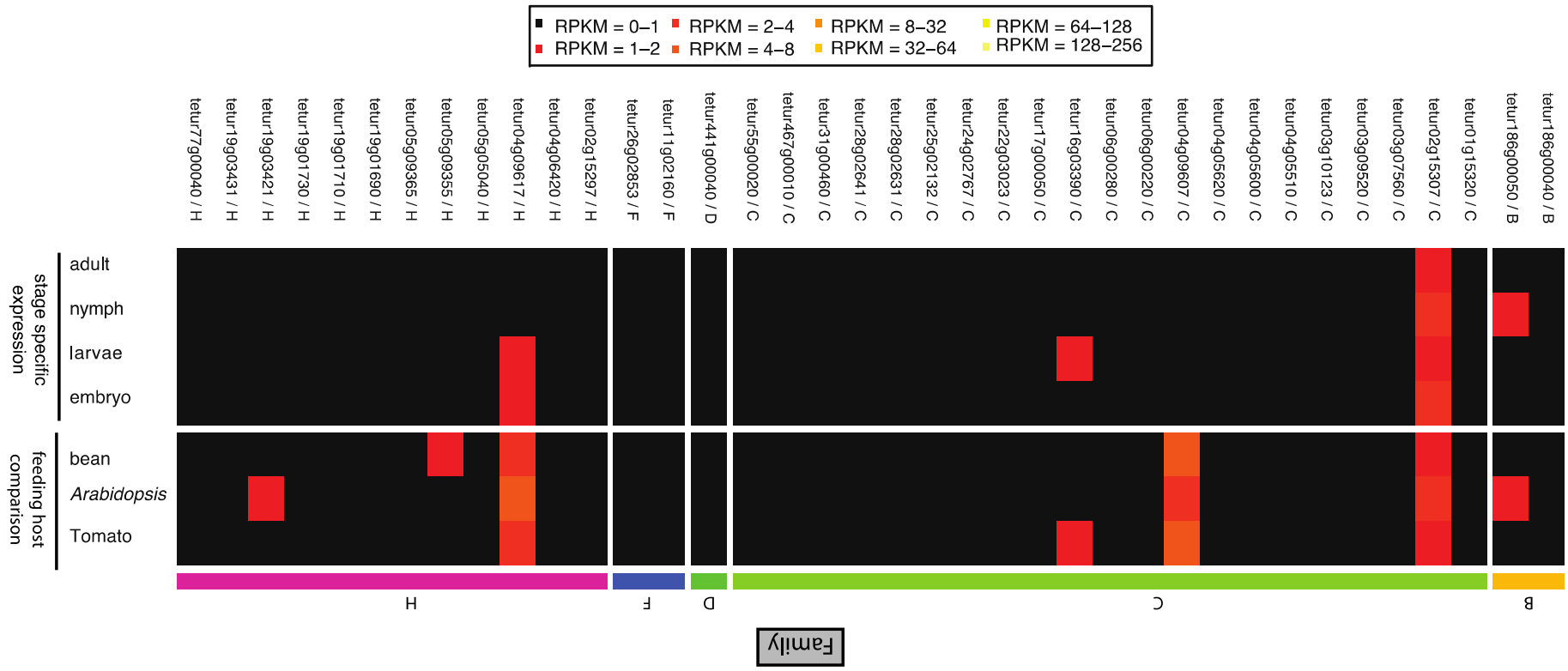


Figure V.10 – Heat plot of mean expression values (RPKM) of *T. urticae* ABC fragments from mites on different host plants (bean, tomato and *Arabidopsis*) and of expression values of *T. urticae* ABC fragments from four different life stages (embryo, larvae, nymph and adult).

Many *T. urticae* ABC genes were broadly expressed, and more than half (57 genes, or 55% of the total) were expressed across all developmental stages analyzed (embryos, larvae, nymphs, and adults). However, the expression of many ABCC and ABCG genes was restricted (or at their highest levels) in larvae and nymphs; overall, embryos and adult females had the highest number of non-expressed ABCs (40 and 36 ABC genes in the adult and embryos, respectively (Fig. V.9)). Furthermore, *tetur30g01400* and *tetur20g02610*, members of the ABCE and ABCF subfamilies respectively, showed very high expression in all stages (Fig. V.9), coinciding with their presumed conserved role in translation regulation (see above). Similar high expression of the *T. castaneum* ABCE gene was reported in all developmental stages examined (Broehan *et al.* 2013). Finally, within the other ABC subfamilies, *tetur27g01890* (ABCA), *tetur11g04030* (ABCB), *tetur04g04360* (ABCC), *tetur05g06640* (ABCD), *tetur02g11270* (ABCG) and *tetur28g00870* (ABCH) had the highest average expression across developmental stages (Fig. V.9).

Twenty-two and 28 ABC genes were differentially expressed in strain London mites transferred from bean to *Arabidopsis* and tomato plants, respectively (Fig. V.11; fold change ≥ 2, FDR < 0.05 as calculated for all *T. urticae* genes, see Methods). We found that 73% (16/22) of the differentially expressed ABC genes in mites transferred to *Arabidopsis* were also differentially expressed on tomato, and belonged to three subfamilies, ABCC (4), -G (5) and -H (7). Surprisingly, ABC genes differentially expressed in two multi-pesticide resistant *T. urticae* strains as detected with expression microarrays (see Chapter IV) belonged to the same ABC subfamilies as identified with RNA-seq data in the host transfer experiment (Table V.4). The ABCC subfamily has been frequently linked with xenobiotic detoxification in arthropods (see above). On the other hand, arthropod members of the ABCG and –H subfamilies have only very recently been reported to be associated with detoxification of xenobiotic compounds (see discussions for the ABCG and ABCH families). However, the differential expression of members from these *T. urticae* ABC gene families (G and H) upon host transfer/exposure to xenobiotics should be further (functionally) validated in following studies.

It is worth noting that the differentially expressed *T. urticae* ABCC, -G and -H genes mentioned above were not among the most highly differentially expressed genes in these experiments, both in expression level and fold change. In total, 893 and 977 differentially expressed genes were identified in the two multi-resistant strains (see Chapter IV).

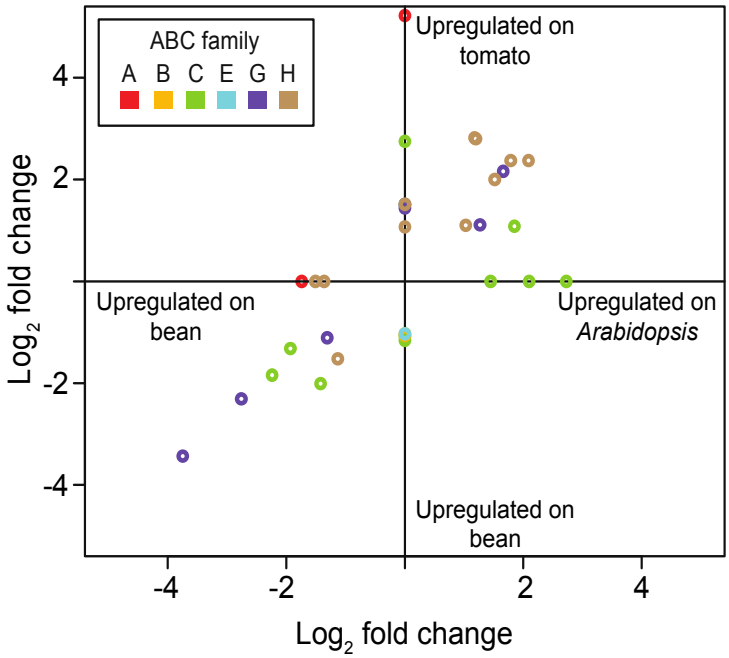


Figure V.11 – Full length differentially expressed (fold change ≥ 2, FDR adjusted p-value < 0.05) ABC genes in mites after host plant change from bean to either *Arabidopsis* or tomato.

Subfamilies are color-coded as follows: ABCA, red; ABCB, yellow; ABCC, green; ABCE, blue; ABCG, purple; and ABCH, brown. Genes found to be significant in only one of the pairwise comparisons have had their fold change values assigned to zero for the non-significant comparison. Fold change values of differentially expressed ABC genes can be found in Appendix V-G.

Table V.4 – Fold changes of full length differentially expressed ABC genes of two multi-resistant strains (MR-VP and MAR-AB) compared to a susceptible strain (London).

Tetur ID	Name	MAR-AB	MR-VP
tetur01g10390	TuABCC-02	2.8731296	
tetur03g07460	TuABCC-07	2.0268726	
tetur03g07490	TuABCC-08	2.0264525	
tetur03g09800	TuABCC-10	2.1275954	
tetur03g09880	TuABCC-11	2.0427277	
tetur04g05540	TuABCC-13		-2.295
tetur05g04300	TuABCC-17	2.4518309	
tetur40g00010	TuABCC-39	2.0048232	
tetur18g00230	TuABCH-13		2.21316
tetur21g00940	TuABCH-15		2.19474

Likewise, 2,502 and 3,951 differentially expressed genes were detected in mites when fed on *Arabidopsis* and tomato in comparison with bean, respectively (Grbić *et al.* 2011). The Major Facilitator Superfamily, another large and widespread transporter family (Pao *et al.* 1998), showed overall a more pronounced response both in the number of genes differentially expressed, and in the fold change values of differentially expressed genes in both resistant strains and after host plant change (see Chapter IV) (Grbić *et al.* 2011). Thus, despite the exceptional number of ABC transporters in the *T. urticae* genome, other transporters and non-transporter proteins also might play key roles in the detoxification of xenobiotics (see Chapter IV) (Grbić *et al.* 2011).

4 Conclusions

The spider mite *T. urticae* is among the most polyphagous pests worldwide and is notorious for its ability to develop resistance against numerous pesticides. One of the prerequisites to study xenobiotic metabolism (pesticides and plant secondary metabolites) in this species is to inventory genes related to detoxification. Here, we provide a survey of the ABC gene superfamily, whose members have frequently been reported to play roles in detoxification, either by directly transporting toxicants out of cells, or after conjugation with glutathione. We identified 103 ABC genes (distributed over eight subfamilies, ABCA-H) in the genome of the spider mite *T. urticae*. To date, this is the largest number of ABC genes reported in any metazoan species. The large number is mainly due to lineage-specific expansions in subfamilies C, G and H. Of particular note, most of the differentially expressed ABC genes in acaricide resistant strains and after introduction of mites to challenging host plants belong to these expanded ABC subfamilies. This hints at their potential role in detoxification and may explain their retention after duplication in the mite genome. However, obtaining functional evidence that members of these ABC subfamilies contribute to xenobiotic tolerance should be the priority of further research.

Due to the lineage specific expansions in the ABCC, G, and H families, inferring the function of specific *T. urticae* ABC family members based on phylogenetic relationships is not straightforward. Nevertheless, we found clear orthologous relationships between some of the *T. urticae* ABC proteins and human ABCC10, ABCG5 and ABCG8, the *D. melanogaster* sulfonylurea receptor and the ecdysone-regulated transporter E23. Furthermore, we found a high conservation between *T. urticae* ABC proteins and members of the ABCB-half transporters and ABCD, -E, and –F subfamilies, which are known to be involved in fundamental processes. To conclude, this study provides the first thorough ABC gene analysis of a polyphagous arthropod herbivore and represents a useful resource for future biochemical and toxicological studies on the role of ABC transporters in the extremely broad host range and development of pesticide resistance of *T. urticae*.

Chapter VI

The cys-loop ligand-gated ion channel gene family of *Tetranychus urticae*: implications for acaricide toxicology and a novel mutation associated with abamectin resistance

---

This chapter has been redrafted from:

Dermauw W., Ilias A., Riga M., Tsagkarakou A., Grbić M., Tirry L., Van Leeuwen T., & Vontas J. (2012). The cys-loop ligand-gated ion channel gene family of *Tetranychus urticae*: implications for acaricide toxicology and a novel mutation associated with abamectin resistance. *Insect Biochemistry and Molecular Biology* 42(7):455-465.

1 Introduction

Members of the cys-loop ligand-gated ion channel (cysLGIC) family are major targets for currently successful insecticides/acaricides (Bloomquist 2003; Buckingham *et al.* 2005; Wolstenholme 2010). However, their identity, phylogeny and physiology are underexplored in chelicerates such as *T. urticae*.

A number of cysLGICs are present in invertebrates, including the cation permeable nicotinic acetylcholine receptors (nAChRs), the γ-aminobutyric acid (GABA)-gated channels, the glutamate-gated chloride channels (GluCls), the histamine-gated chloride channels (HisCls), and the pH sensitive chloride channels (pHCl). Each cysLGIC consists of five subunits which assemble to form binding sites for activating neurotransmitters and a selective ion pore that crosses the cell membrane. A universal feature of each cysLGIC subunit is a common topology of four membrane-spanning segments (TM1–TM4) and a large N-terminal extracellular domain with a conserved cysteine-bridge motif (Schnizler *et al*. 2005). cysLGICs are distributed throughout the nervous system of arthropods and are involved in a wide range of biological functions, including synaptic inhibition, cellular excitability, pH regulation and organic solute transport, by mediating inhibitory and excitatory responses to neurotransmitters (Ortells and Lunt 1995).

The cysLGIC gene family is relatively small in arthropods with 21-26 members in insect genomes analysed to-date (Jones and Sattelle 2006; Jones and Sattelle 2007; Dale *et al.* 2010; Jones *et al.* 2010; Knipple and Soderlund 2010), compared to the much more expanded families present in the nematode *Caenorhabditis elegans* and in vertebrates (Jones and Sattelle 2008; Jones and Sattelle 2010). Structural diversity in arthropods (as discovered in *Drosophila*) is achieved by post-transcriptional and/or post-translational modifications and has been suggested to compensate for the small number of genes (Knipple and Soderlund 2010). With the exception of a few cysLGIC gene fragments (Lees and Bowman 2007; Mounsey *et al.* 2007; Kwon *et al.* 2010) that have been sequenced and deposited in the database, this gene family has not been studied in the subphylum of the Chelicerata, the second largest group of terrestrial animals and a basal branch of arthropods.

Members of the cysLGIC are known to mediate the toxicity of a variety of insecticides. Neonicotinoid insecticides and spinosad have high and selective affinity for many arthropod nAChRs (Jeschke and Nauen, 2008). Cyclodienes (such as dieldrin) and phenylpyrazoles (such as fipronil) mainly target GABA receptors (Buckingham *et al.* 2005). The macrocyclic lactones, the avermectins and the ivermectins have high affinity for GluCls (Wolstenholme 2010), which are only found in invertebrates and in all of the avermectin targeted pests (Kehoe *et al.* 2009). The toxicity of some of the aforementioned insecticides is either very limited (i.e. neonicotinoids) or not well established (i.e. phenylpyrazoles) in the spider mite *T. urticae*. In contrast, avermectins (such as abamectin), are very efficient for the control of *T. urticae*, including widely distributed insecticide/acaricide resistant populations around the world. However, resistance of *T. urticae* against abamectin has been reported (Kwon *et al*. 2010; Van Leeuwen *et al*. 2010a). By comparing sequences of a GluCl subunit (Tu_GluCl1) between isogenic abamectin resistant and susceptible strains, Kwon *et al*. (2010) identified a G323D substitution which was tightly associated with a moderate resistance phenotype (17-fold resistance).

The recently sequenced *T. urticae* genome, the first amongst Chelicerates (Grbić *et al.* 2011), has provided a unique opportunity to study the complete set of genes encoding spider mite cysLGIC subunits and understand their role in mite biology and their interaction with acaricides. Here, we annotate the cysLGIC gene family of *T. urticae* and analyse its phylogenetic relationships with insects. Toxicological implications associated with specific features of these receptors in *T. urticae* compared to insects are discussed. Given the expansion of GluCl genes in *T. urticae*, we investigate in more detail their putative role in abamectin resistance and reveal that an accumulation of mutations in different genes, including a novel one that has not been previously reported, are associated with high levels of abamectin resistance.

2 Materials and methods

2.1 Identification of cysLGIC subunits in the *T. urticae* genome and phylogenetic analysis

To identify putative cysLGIC genes, a tBLASTn-search (E-value $>10^{-4}$) was conducted against the *T. urticae* genome (version of November 2011, http://bioinformatics.psb.ugent.be/orcae/overview/Tetur) using protein sequences of the LGIC family of *Drosophila melanogaster* as queries (Jones *et al.* 2010). Candidate *T. urticae* genes were identified based on their sequence homology with previously characterized subunits, particularly in the N-terminal ligand-binding domain and the four transmembrane regions. Partial fragments (Table VI.1) of cysLGIC genes were separated from putative full-length cysLGIC genes and gene models (predicted by the Eugene-platform, (Grbić *et al.* 2011)) were refined on the basis of homology, multiple alignments and available EST support. Signal-peptide cleavage sites were predicted using SignalP 4.0 (Petersen *et al.* 2011) with a D-cutoff value of 0.45 and membrane spanning regions were identified by alignment with *D. melanogaster* representative cysLGIC subunit protein sequences.

The final set of *T. urticae* cysLGIC subunit protein sequences were aligned with those of *D. melanogaster*, *Tribolium castaneum*, *Apis mellifera*, *Rhipicephalus microplus* (see Table VI.2 for GenBank accession numbers) and a representative set of mammalian glycine gated chloride channels protein sequences using MUSCLE version 3.8.31(Edgar 2004). Gene names were assigned following the closest *Drosophila* orthologue. According to the Akaike information criterion, the WAG+I+G+F model was optimal for phylogenetic analysis. A maximum likelihood analysis was performed using Treefinder (Jobb *et al.* 2004), and bootstrapped with 1000 pseudoreplicates.

Table VI.1 - List of cysLGIC gene fragments in *T. urticae*

tetur ID	length (bp)	best blastp-hit in *T. urticae* proteome
tetur125g00010	180	tetur27g02270 (Tuα7)
tetur289g00010	330	tetur09g00790 (Tuα4)
tetur692g00020	207	tetur09g00750 (Tuα2)

Table VI.2 - Accession numbers of cysLGIC subunit protein sequences*

T. castaneum		*D. melanogaster*		*A. mellifera*		*T. urticae***	
cysLGIC name	GenBank ID	cysLGIC name	GenBank ID	cysLGIC name	GenBank ID	cysLGIC name	tetur (ORCAE)
Tcasα1	EF526080	Dα1	CAA30172	Amelα1	NP_001091690.1	Tu_12344.1	tetur01g10260
Tcasα2	EF526081	Dα2	CAA36517	Amelα2	NP_001011625.1	Tu_12344.2	tetur02g11020
Tcasα3	EF526082	Dα3	CAA75688	Amelα3	NP_001073029.1	Tu_12344.3	tetur02g11170
Tcasα4	EF526083	Dα4	CAB77445	Amelα4	NP_001091691.1	Tu_12344.4	tetur11g04710
Tcasα5	EF526085	Dα5	AAM13390	Amelα5	XP_625040.1	Tu_GluCl1	tetur02g04080
Tcasα6	EF526086	Dα6	AAM13392	Amelα6	NP_001073564.1	Tu_GluCl2	tetur08g04990
Tcasα7	EF526089	Dα7	AAK67257	Amelα7	NP_001011621.1	Tu_GluCl3	tetur10g03090
Tcasα8	EF526090	Dβ1	CAA27641	Amelα8	NP_001011575.1	Tu_GluCl4	tetur22g02450
Tcasα9	EF526091	Dβ2	CAA39211	Amelα9	NP_001091694.1	Tu_GluCl5	tetur36g00090
Tcasα10	EF526092	Dβ3	CAC48166	Amelβ1	NP_001073028.1	Tu_GluCl6	tetur41g00120
Tcasα11	EF526093	GluCl	AAG40735	Amelβ2	NP_001091699.1	Tu_HisCl1	tetur25g00320
Tcasβ1	EF526094	Grd	Q24352	Amel_GluCl	NP_001071277.1	Tu_HisCl2	tetur09g01700
Tcas_Clgc1	EF545129	HisCl1	AAL74413	Amel_Grd	XP_397242.4	Tu_HisCl3	tetur15g02290
Tcas_Clgc2	EF545130	HisCl2	AAL74414	Amel_HisCl1	NP_001071279.1	Tu_HisCl4	tetur17g02170
Tcas_Clgc3	EF545131	Lcch3	AAB27090	Amel_HisCl2	ABG75740.1	Tu_LGIC_unk	tetur41g00670
Tcas_GluCl	EF545121	Ntr	AF045471	Amel_Lcch3	NP_001071280.1	Tu_pHCl	tetur07g04510
Tcas_Grd	EF545119	pHCl	NP_001034025	Amel_pHCl	NP_001071284.1	Tu_Rdl1	tetur12g03620
Tcas_HisCl1	EF545124	Rdl	AAA28556	Amel_Rdl	XP_001120292.2	Tu_Rdl2	tetur36g00580
Tcas_HisCl2	EF545125	CG6927	AAF45992	Amel_6927	ABG75747.1	Tu_Rdl3	tetur36g00590
Tcas_Lcch3	EF545120	CG7589	AAF49337	Amel_8916	NP_001071290.1	Tuα1	tetur09g00760
Tcas_pHCl	EF545126	CG8916	BT022901	Amel_12344	XP_623445.2	Tuα2	tetur09g00750
Tcas_Rdl	EF545117	CG11340	AAF57144			Tuα3	tetur01g10770
Tcas_8916	EF545127	CG12344	AAF58743			Tuα4	tetur09g00790
Tcas_12344	EF545128					Tuα5	tetur04g07390
						Tuα6	tetur27g02250
						Tuα7	tetur27g02270
						Tuβ1	tetur09g00800
						Tuβ2	tetur08g07620
						Tuβ3	tetur08g07630

R. microplus		*C. elegans*		*M. persicae*		*P. pseudoannulata*	
cysLGIC name	GenBank ID	cysLGIC name	GenBank ID	cysLGIC name	GenBank ID	cysLGIC name	GenBank ID
Rmic_GABA	ACV07674.1	CeGlc-3	NP_504441.1	Mperβ1	CAB87995.1	Ppseβ1	ADG63463
		CeGAB-1	NP_499661				

I. scapularis		*N. lugens*	
cysLGIC name	GenBank ID	cysLGIC name	GenBank ID
Iscaβ1	XP_002406474	Nlugβ1	ACJ07013
Isca_GABA	XP_002411566.1		

Vertebrates		
species	cysLGIC name	GenBank ID
Homo sapiens	Hs_GlyRα2	BAH12535
Danio rerio	Dr_GlyRα2	NP_001161371
Bos taurus	Bt_GlyRα2	NP_001179843
Equus caballus	Ec_GlyRα2	XP_001489514
Sus scrofa	Ss_GlyRα2	XP_001926528
Callithrix jacchus	Cj_GlyRα2	XP_002762681

* accession numbers higlighted in grey were used for phylogenetic analysis
** accesion numbers of *T. urticae* can be downloaded from the ORCAE website (http://bioinformatics.psb.ugent.be/orcae/overview/Tetur)

2.2 Mite strains

The London reference strain originates from a wild-collected *T. urticae* population from the Vineland region (Ontario, Canada) and DNA from an inbred line of this strain was used for *T. uticae* genome sequencing (Grbić *et al.* 2011). The MAR-AB strain was isolated from a heavily sprayed rose greenhouse near Athens in 2009, and was maintained under abamectin selection (10 mg/L abamectin) every two generations, until homogeneity in bioassay response was obtained. Strain 033 and LS-VL were previously described (Van Leeuwen *et al.* 2004; Khajehali *et al.* 2011). *T. urticae* strains were mass reared on potted kidney bean plants at 26°C (±0.5°C), 60% relative humidity (RH) and 16/8 h light/dark photoperiod.

2.3 Extraction of genomic DNA and RNA, cDNA synthesis, cloning, and sequencing

Genomic DNA (gDNA) was extracted from individual mites as previously described (Van Leeuwen *et al.* 2008). Total RNA was extracted from about 100 adult female mites using RNeasy Mini kit (Qiagen). First strand cDNA was synthesized from 2.5 µg of total RNA using Maxima first strand cDNA synthesis kit (Fermentas) in a volume of 20 µl. Primer pairs for all RT-PCR amplifications and single mite genotyping were designed using Primer 3 (Rozen and Skaletsky 2000) (see Table VI.3, primers with "dia'' suffix were used for genotyping). *T. urticae* GABA- and glutamate- gated chloride channels were amplified in 2 to 4 overlapping gene fragments from 0.5 µl cDNA in 1 x Taq reaction buffer (Invitrogen, Belgium), with a final concentration of 2 mM MgCl2, and 0.2 µM of each primer and cycling conditions of 92°C for 2 min, 37 cycles of 10 s at 92°C, 30 s at 55°C, 1 min at 72°C; and 2 min at 72°C. PCR-products were purified using E.Z.N.A. Cycle Pure Kit (Omega Bio-Tek) and were either directly sequenced (LGC genomics) or ligated into the pJET1.2 vector (Fermentas) and sequenced with pJET1.2F and pJET1.2R primers (Table VI.3). GABA- and glutamate-gated chloride channel sequences of the MAR-AB strain were submitted in the GenBank database (GenBank accession numbers: JQ738191- JQ738198)

2.4 Sequence analysis

All sequence data were analysed using BioEdit version 7.0.1 (Hall 1999). Identity and similarity matrices of *T. urticae* cysLGIC subunit protein sequences with *D. melanogaster* cysLGIC subunits were calculated using MatGAT 2.0 (Campanella *et al.* 2003). Hierarchical clustering (with average linkage) of RNA-sequence data of cysLGIC genes from four

different life stages (egg/embryo, larva, nymph and adult) of *T. urticae* (Grbić *et al.* 2011) was performed using GENESIS (Sturn *et al.* 2002).

Table VI.3 - Primers used in this study

Tetur-ID	Primer Name	Sequence (5'-3')	T_a (°C)[a]	Product size (bp)
tetur12g03620	Tu_Rdl1aF	TTGTTAAAGCCAATCTCATTGC	55	701
	Tu_Rdl1aR	GCAATTGAGAACTGAGGAAGC		
tetur12g03620	Tu_Rdl1bF	CGGTAACTGCAAGTTGTCCA	55	761
	Tu_Rdl1bR	CATTGGATTGGTTGTTGTGC		
tetur12g03620	Tu_Rdl1cF	GTATGCGGCTGTTGGTTACA	55	893
	Tu_Rdl1cR	CTACACCTGAACTGGCAGCA		
tetur12g03620	Tu_Rdl1dF	TGGAGGTAGTGGGAACCTCA	55	712
	Tu_Rdl1dR	CCAACTGTGACTCTTGCCTTT		
tetur36g00580	Tu_Rdl2aF	TCTCATTTTTAACTCAATGAGGAAG	55	637
	Tu_Rdl2aR	AGCAGCTATGGCTGATTCTTG		
tetur36g00580	Tu_Rdl2bF	TGGATAGACAAGCCTGCTCA	55	769
	Tu_Rdl2bR	TTCATTTTCTGAGGGCAACC		
tetur36g00580	Tu_Rdl2cF	TGCTGTAGCCGAACAAAGTG	55	510
	Tu_Rdl2cR	TTTTGAGTCTTTCGCCTGAA		
tetur36g00590	Tu_Rdl3aF	TTGATGTCGGTGACAAAGGA	55	1,027
	Tu_Rdl3aR	CGTTTCTTTTGGTCTTCATGC		
tetur36g00590	Tu_Rdl3bF	GCCAATGCTTTAAGTACAGGAAA	55	910
	Tu_Rdl3bR	GCCTTGGATTACGTGCATTC		
tetur36g00590	Tu_Rdl3cF	GGTGGACAAGGTGGACAAAT	55	846
	Tu_Rdl3cR	ATCTGGATTGGGTTCATTGC		
tetur02g04080	Tu_GluCl1a	CAATATGAGTGATCCATTGACGA	55	785
	Tu_GluCl1a	CCAACATACAACAGGGCACA		
tetur02g04080	Tu_GluCl1b	AACCAAATCGCTTCACTTGC	55	717
	Tu_GluCl1b	TCCATAAGCTCATCTCGAAGAA		
tetur02g04080	Tu_GluCl1_dia_F	TTGGATTGACCCTAACTCAGCA	54	263
	Tu_GluCl1_dia_R	TTGCACCAACAATTCCTTGA		
tetur08g04990	Tu_GluCl2a	TGACTTTGGCTGTTTTCACG	55	786
	Tu_GluCl2a	TGAACTCTCTTTTAAAGTCAAGGTACA		
tetur08g04990	Tu_GluCl2b	CCGATGATCTGGAGTTTCGT	55	781
	Tu_GluCl2b	CGTCATCGTGATGCATCTTT		
tetur08g04990	Tu_GluCl2_dia_F	TCATCGTCTCTTGGGTCTCC	54	282
	Tu_GluCl2_dia_R	CCCATCGTCGTTGATACCTT		
tetur10g03090	Tu_GluCl3a	GTTGTGTCTACCTGGACCAAAGT	55	730
	Tu_GluCl3a	TTCACCTGTATTTGTTCGTGAAG		
tetur10g03090	Tu_GluCl3b	TCCTGCCCAATGAATTTGA	55	844
	Tu_GluCl3b	GGAGTTAACCCAAAATCACCA		
tetur10g03090	Tu_GluCl3_dia_F	CCGGGTCAGTCTTGGTGTTA	54	251
	Tu_GluCl3_dia_R	CACCACCAAGAACCTGTTGA		
tetur22g02450	Tu_GluCl4a	TCAATCAACCAGCATCCTACC	55	788
	Tu_GluCl4a	AATGTAAGTCTAACTACGATGCAACTG		
tetur22g02450	Tu_GluCl4b	CATCTTCCTCGATTCGGTTT	55	800
	Tu_GluCl4b	TGCGTTTTTGTGTGCTTCTT		
tetur22g02450	Tu_GluCl4_diaF	TATTCCAGCCCGAGTTTCAC	54	250
	Tu_GluCl4_diaR	AATCGGAGGTTGACTTGGTG		
tetur36g00090	Tu_GluCl5a	AGCTTAAGAACTATCGAGAGACGC	55	724
	Tu_GluCl5a	AGAATGAAACCCACGAAACG		
tetur36g00090	Tu_GluCl5b	TGTACCATCTCAATGGCATCAT	55	688
	Tu_GluCl5b	CACTGGTTTGGAGTTCATCG		
tetur36g00090	Tu_GluCl5_dia_F	ATGTTGGTCATCGTTTCGTG	54	294
	Tu_GluCl5_dia_R	AATCGGGATTGAATTTGCTG		
	pJET1.2F	CGACTCACTATAGGGAGAGCGGC	53	-
	pJET1.2R	AAGAACATCGATTTTCCATGGCAG		

[a] T_a = annealing temperature

2.5 Toxicity tests

Bioassays were performed on bean leaves by using a commercial formulation of abamectin (Vertimec, 18 g/l 1.8 EC), as previously described (Tsagkarakou *et al.* 2009). Briefly, individual Petri dishes with leaf discs on wet cotton wool were sprayed with a precision Potter Spray Tower (Burkard, Rickmansworth, Hertfordshire, UK) (Potter 1952) with 1 ml aqueous solutions of insecticide at 1 bar pressure in order to obtain a homogenous spray film (2.4 mg ± 0.3 aqueous acaricide deposit/cm^2). Bioassay data were analysed using PoloPC (LeOra Software, Berkeley, CA). The toxicity of 200 ppm fipronil to four different strains of *T. urticae* (O33, MAR-AB, LS-VL and London) was assessed using the same bioassay as for abamectin.

2.6 Crossing experiments and single mite genotyping

T. urticae reproduces by arrhenotokous parthenogenesis, meaning haploid males will develop from unfertilized eggs. To assess the dominance of resistance, and estimate the number of genes involved, reciprocal crosses were made by mass matings between susceptible (London) and resistant parents (MAR-AB) through two consecutive generations to produce hybrid F1 and F2 females as previously described (Van Leeuwen *et al.* 2004). The resulting F1 and F2 females were collected and after maturation (1–3 days) their mortality was scored in a bioassay with the appropriate concentrations of abamectin, as described above. Based on LC_{50} values, the degree of dominance was determined with the formula of Stone (Stone 1968). To test for monogenic resistance, the observed response in bioassays was compared to the expected response with the formula: C=0.5 W (parent 1) + 0.5 w (parent 2), where C is the expected mortality and w the observed mortality of the parental types at a given concentration (Georghiou 1969). A χ^2goodness-of-fit analysis was used to examine any deviation between the observed response and the expected response under monogenic resistance. Single mite genotyping of segregating F2 of both crosses was done as previously described (Van Leeuwen *et al.* 2008) by using gene specific primers (Table VI.3 primers with "dia" suffix) on both F2 progenies, ((SS♀ x R♂) x S♂) and ((RR♀ x S♂) x R♂), before and after selection with 10 and 100 mg/L abamectin, respectively. A χ^2 test was performed to test the absence of linkage disequilibrium between mutations.

3 Results and Discussion

3.1 The spider mite cysLGIC gene superfamily

We have characterized the cysLGIC gene superfamily of the spider mite *T. urticae*, the first one within the subphylum of the Chelicerata. Using tBLASTn and manual annotation, we identified 29 *T. urticae* cysLGIC subunits with conserved superfamily features, including 1) an extracellular N-terminal domain with distinct regions (loops A-F) forming the ligand binding site 2) the dicysteine-loop 3) four transmembrane regions (TM1-4) and 4) a highly variable loop between TM3 and TM4 (Jones *et al.* 2010) (see Appendix VI-A and B). The spidermite cysLGIC gene superfamily represents the largest one among arthropods described to date (Table VI.4), since the honey bee (*A. mellifera*), the pea aphid *Acyrthosiphon pisum*, the fruit fly (*D.melanogaster*), the flour beetle (*T. castaneum*) and the parasitoid jewel wasp (*Nasonia vitripennis*) have 21, 21, 23, 24 and 26 genes, respectively (Table VI.4) (Jones and Sattelle 2006; Jones and Sattelle 2007; Dale *et al.* 2010; Jones *et al.* 2010; Knipple and Soderlund 2010). A maximum likelihood phylogenetic analysis of *T. urticae, D. melanogaster*, *A. mellifera, T. castaneum, R. microplus* cysLGIC subunits and representative mammalian glycine gated chloride channels protein sequences, identified *T. urticae* cysLGIC subunit families with high bootstrap support, and indicated orthologous relationships between spider mite and insect subunits in many cases (Fig. VI.1). RNA-seq profiling data from an earlier study (Grbić *et al.* 2011) showed that all but one (*Tu_GluCl5*) *T. urticae* cysLGIC gene are transcribed in at least one life stage (Fig. VI.2). A more detailed description of the cysLGIC gene subgroups and their relationships with other species is presented below.

3.2 *T. urticae* nicotinic acetylcholine receptor subunits

The nAChRs are highly abundant ligand-gated cation channels that are activated by acetylcholine and mediate fast cholinergic synaptic transmission in arthropods and vertebrate nervous systems (Jones and Sattelle 2010). They consist of five homologous subunits arranged around a central ion channel. We identified 10 nAChR genes (Fig. VI.1 and Appendix VI-A) in the *T. urticae* genome, which is comparable to the number of genes present in insect genomes sequenced to date (e.g. *D. melanogaster*, *T. castaneum*, *A. mellifera* and *N. vitripennis* have 10, 12, 11 and 16 nAChR genes, respectively) (Table VI.4,

(Jones and Sattelle 2006; Jones and Sattelle 2007; Jones *et al.* 2010; Knipple and Soderlund 2010)).

Table VI.4 - cysLGIC genes of *D. melanogaster* (Dm)), *A. mellifera* (Am), *T. confusum* Tc, *N. vitripennis* (Nv) and *T. urticae* (Tu)

cysLGIC gene group[a]	Dm	Am	Tc	Nv	Tu
nicotinic acetylcholine receptors (nAChRs)					
Dα1 group	1	1	1	1	0
Dα2 group	1	1	1	1	1
Dα3 group	1	1	1	1	0
Dα4 group	1	1	1	1	0
Dα5-7 group	3	3	3	3	3
Dβ1 group	1	1	1	1	1
Dβ2 group	1	1	2	1	1
divergent nAChRs	1	2	2	7	2
unknown	0	0	0	0	2
GABA -gated ion channels					
Rdl	1	1	1	1	3
Grd	1	1	1	1	0
Lcch3	1	1	1	1	0
other cysLGICs					
CG8916	1	1	1	1	0
Insect group 1 of cysLGIC	3	1	3	1	0
histamine-gated chloride channels	2	2	2	2	4
glutamate-gated chloride channels	1	1	1	1	6
pH-sensitive chloride channels	1	1	1	1	1
unknown group	0	0	0	0	1
group 12344 of cysLGIC	1	1	1	1	4
Ntr	1	0	0	0	0
total number of cysLGIC genes	23	21	24	26	**29**

[a] group names according to Jones *et al.* 2010, numbers are derived from (Jones *et al.*, 2010; Jones and Sattelle, 2006; Jones and Sattelle, 2007; Knipple and Soderlund, 2010) and this study

Seven of the *T. urticae* nAChR subunits possess the two adjacent cysteine residues in loop C which defines them as *α* (alpha) subunits, while three lack the vicinal cysteine residues and are thus designated as β (beta) subunits (Appendix VI-A). We detected the GEK-motif preceding transmembrane region 2 in all *T. urticae* nAChR subunits (Appendix VI-A), except for Tuβ1 and Tuβ2. This GEK-motif is thought to play an important role in ion permeability and selectivity (Jensen *et al.* 2005). Remarkably, Tuα3 has an atypical FxCC amino-acid motif (Appendix VI-A), instead of the highly conserved YxCC in loop C. A study of the chicken α7 nAChR receptor expressed in *Xenopus* oocytes showed that changing the YxCC motif to FxCC, resulted in a tenfold decrease in ACh affinity (Galzi *et al.* 1991). The unusual FxCC motif was previously also identified in nAChR(s) subunits of *Tribolium*

confusum, N. vitripennis and the nematodes *C. elegans* and *Haemonchus contortus* (Jones *et al.* 2010; Rufener *et al.* 2010). Future functional studies should point out if the presence of this motif in arthropod and nematode nAChRs also results into a similar decrease in ACh affinity.

A phylogenetic analysis revealed six *T. urticae* orthologues (Tuα2, Tuα4, Tuβ1, Tuα5, Tuα6 and Tuα7, Fig. VI.1) of insect nAChRs subunits. Insects possess at least one divergent subunit that shares low sequence identity with other nAChR subunits (Jones *et al.* 2010). Similar to insects, we identified two divergent *T. urticae* nAChR subunits (Tuβ2 and Tuβ3, Fig. VI.1), which are located next to each other on scaffold 8 and share high amino acid sequence identity (71%) but low identity with other nAChRs (Table VI.5).

Table VI.5 - Percentage identity/similarity between *T. urticae* and *D. melanogaster* nAChR subunit protein sequences

	Tuα5	Tuβ2	Tuβ3	Tuα2	Tuα1	Tuα4	Tuβ1	Tuα6	Tuα3	Tuα7
Dα1	30.9/52.8	20.6/38.3	20.1/38.4	48.1/67.4	50.8/67.3	48.8/70.4	33.9/50.5	30/54.1	27.2/49.6	30.6/50.6
Dα2	29.3/52.8	21/40.3	19.9/37.5	48.4/67.2	44.8/64.7	45/66.1	32.6/49.6	30.5/53.5	29.3/51.5	30/50.9
Dα3	27.3/41.8	16.3/29.2	15.8/28.2	33.3/46.7	37.4/53.7	37.8/51.2	33.8/53.1	24.9/41.6	26.2/45.5	28.5/44.5
Dα4	30.7/54.2	20.3/37.5	18.7/37.7	45.6/64.4	45.6/65.6	48.7/67.1	33.4/51.8	31.4/53.7	28.9/49.6	29.5/50.9
Dα5	31.9/48.1	14.4/26.6	14.9/26.1	26.1/39.7	25.6/42.1	24.1/39.3	23.2/43.1	31.8/47.1	24.3/43.4	31.1/48.3
Dα6	46.6/62.1	23/41.5	23.1/41.1	34.7/52.5	34.6/51.5	36/55.1	28.9/45.3	45.5/65.9	30.9/48.8	45.8/59.7
Dα7	17.5/24.8	13.1/24.6	13/23.7	14.8/23.4	14.3/21.8	12.3/21.9	11.1/19.5	18.6/27.1	14.5/20.1	17.7/23.4
Dβ1	30.2/51.5	22.5/43.4	21.1/42	34.7/55	38.5/56.8	40.5/61.9	51.8/59.4	30.2/57.3	27.4/46.2	29.5/49.2
Dβ2	27.4/50.1	20/43.2	20.9/40	45.3/64.6	43.2/61.4	49.6/67.5	30/47	29.5/54.7	27.2/47.7	27.1/47.2
Dβ3	17.4/36.1	25/45.6	23/45.1	20.4/41.5	18.6/37.5	21.3/41.2	17.8/32.4	20.6/40.9	21/37.4	18.9/35.2
Tuα5		19.4/36.8	19/37	28.7/50.3	31/51.8	30.6/50.8	29.5/48.2	45.1/64.4	32.4/51	49/64
Tuβ2			71.2/81.6	21.5/40.8	21.2/37.5	21.4/40.4	16.9/32.7	22.6/38.6	19.3/35.6	19.8/34
Tuβ3				21.7/39.2	19.9/35.3	22.9/39.2	17.9/33.5	23/40.1	19.4/36.3	19.3/35.2
Tuα2					41.9/60.3	46.1/66.2	29.9/47.1	30.9/54.1	27.5/48.4	29/48.4
Tuα1						44.9/61.7	35.2/53.9	32.4/51.3	29.4/53.2	33.3/54.1
Tuα4							33.6/49.2	33/56.1	28.7/47.4	31.4/50.5
Tuβ1								26.2/45	26.5/48.8	30.2/49.6
Tuα6									32.2/49.8	44.1/62.6
Tuα3										30.9/53.8
Tuα7										

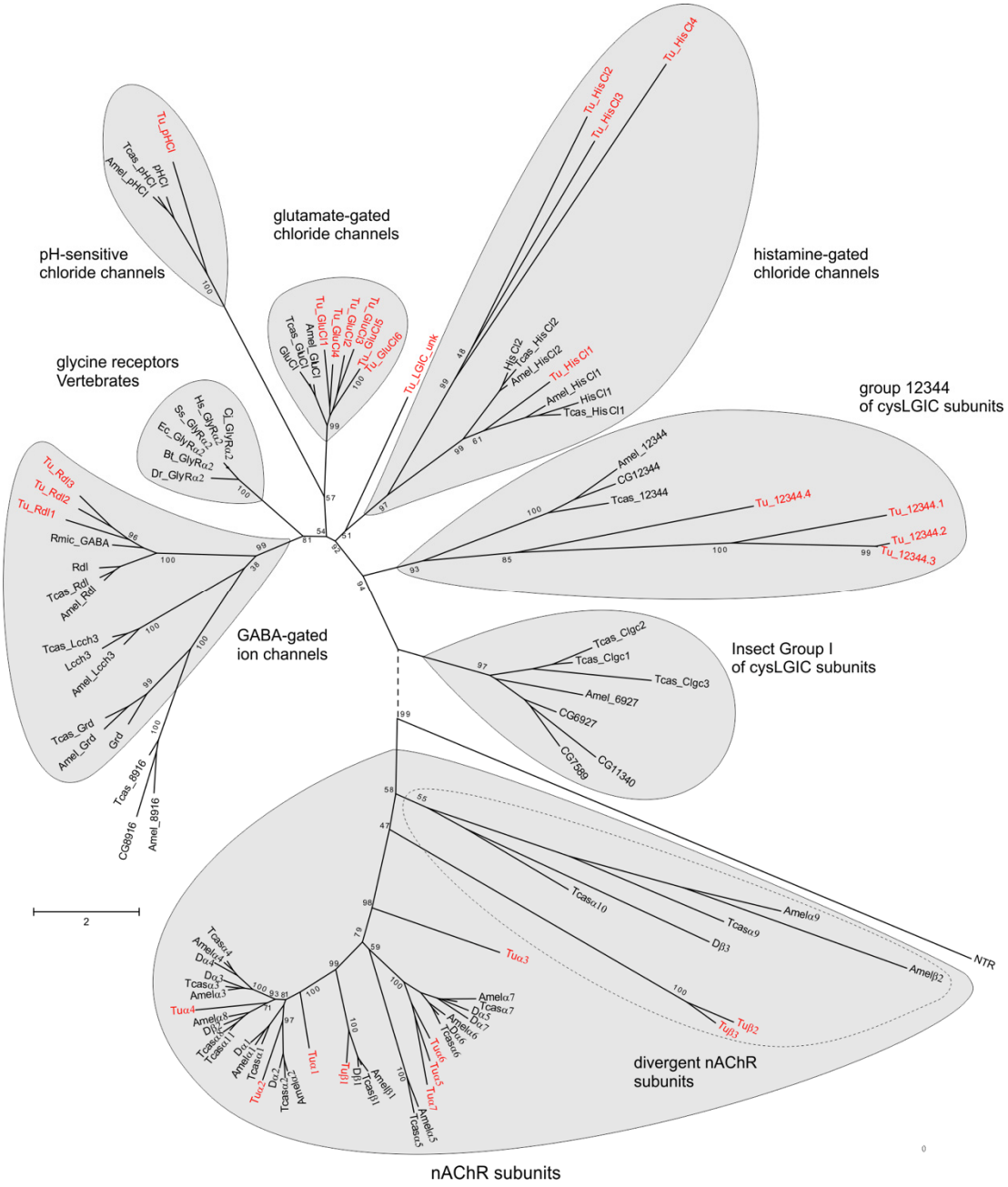


Figure VI.1 - Maximum likelihood phylogenetic tree of *T. urticae* (Tu, indicated in red), *D. melanogaster*, *A. mellifera* (Amel), *T. castaneum* (Tcas), *R. microplus* (Rm) and vertebrate cysLGIC subunit protein sequences.

Sequences were aligned using MUSCLE and a bootstrapped unrooted tree was constructed. Bootstrap values are indicated for crucial branches. For accession numbers see Table VI.2.

Chamaon *et al*. (2002) suggested that gene clustering may favor coordinated expression and co-assembly of nAChR subunits. In *T. urticae, Tuα1*, *Tuα2*, *Tuα4* and *Tuβ1* are located next to each other on scaffold 9, *Tuβ2* and *Tuβ3* are found aside each other on scaffold 8, while *Tuα6* and *Tuα7* are tightly clustered (within 7 kb of each other) on scaffold 27. Hierarchical clustering analysis of *T. urticae* nAChR expression RNA-seq data (Fig. VI.2) groups the expression pattern of *Tuα1*, *Tuα4* and *Tuβ1*, suggesting that these three subunits might be assembled together into one specific nAChR subtype. Nicotinic receptors are the targets of neonicotinoids, one of the most successful insecticide classes in the market (Jeschke and Nauen 2008). However, neonicotinoids are not particularly active against spider mites (James and Price 2002; Ako *et al.* 2004). Several features of insect nAChR subunits have been proposed to contribute to neonicotinoid selectivity and resistance in insects (Fig. VI.3) (Shimomura *et al.* 2004; Liu *et al.* 2005; Shimomura *et al.* 2006; Yao *et al.* 2008; Song *et al.* 2009; Bass *et al.* 2011), and some of the features associated with resistant or insensitive insects are indeed present in *T. urticae*. Firstly, an insertion in the loop B-C interval. In *D. melanogaster* 5 nAChR subunits (*Dα1, Dα1, Dα3, Dα4* and *Dβ2*) have an insertion in the loop B-C interval, which may interact with the neonicotinoid imidacloprid (Shimomura *et al.* 2004). However, only three nAChRs of *T. urticae* have this insertion (Tuα1, 2 and 4, Appendix VI-A). Secondly, nAChR insensitive forms: Song *et al*. (2009) showed by functional expression of chimeric receptors that the selective toxicity of imidacloprid between *Myzus persicae* and its predator *Pardosa pseudoannulata* (wolf spider) is due to amino acid differences (R81Q, N137G and F190W, *M. persicae* numbering) in loop D, E and F of nAChR β1. The importance of position 81 in loop D was recently reinforced by a study of Bass *et al*. (2011) that illustrated that a R81T change is strongly associated with neonicotinoid resistance in *M. persicae*. *T. urticae* has similar amino acids (Tuβ1: 81E and 190W in loop D and loop F, respectively) with those present in insensitive forms of arthropod nAChRs (Fig. VI.3). Other positions, such as V83, R135, Y151S, W197 that have been implicated in imidacloprid selectivity (Liu *et al.* 2005; Shimomura *et al.* 2006; Yao *et al.* 2008), are variable in *T. urticae* nAChRs (Fig. VI.3, Appendix VI-A). Thus, it is likely that the described polymorphism of *T. urticae* nAChR subunits is contributing to the spider mite insensitivity to neonicotinoids (Ako *et al.* 2004). However, differential uptake and/or metabolism of neonicotinoids cannot be excluded. Interestingly, similar amino acids with those found in spider mites and resistant insects are also present in the nAChR *β1* receptor of the tick *Ixodes scapularis* (Fig. VI.3), which is also relatively insensitive to neonicotinoids (McCall *et al.* 2004; Baynes 2009).

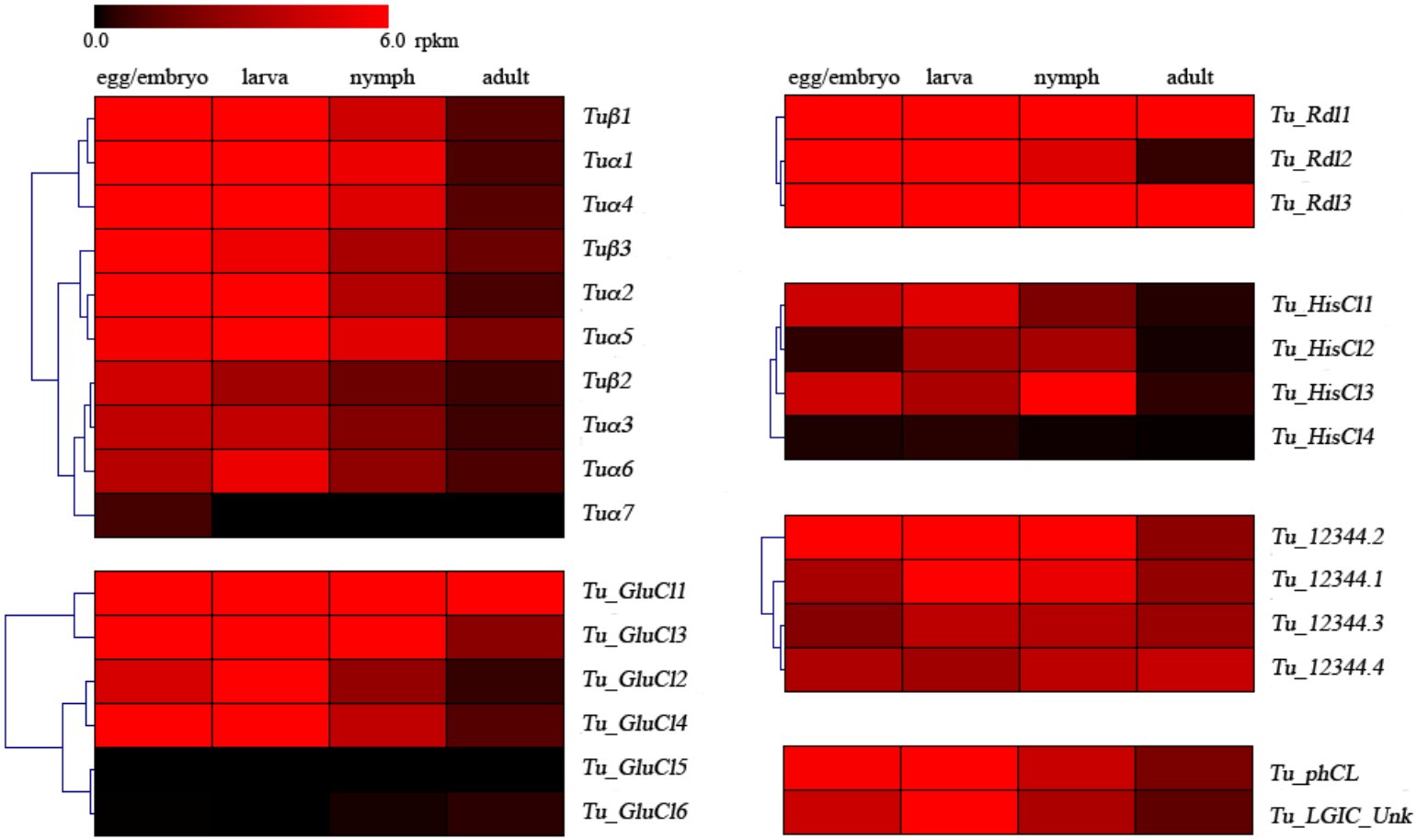


Figure VI.2 - Heat-map of hierarchical clustering analysis (with average linkage) of RNA-seq profiling data of cysLGIC subunits of different life stages (egg/embryo, larva, nymph and adult) of *T. urticae*.

The color bar with corresponding rpkm (reads per kilobase of exon model per million mapped reads) values is shown at the top of the figure.

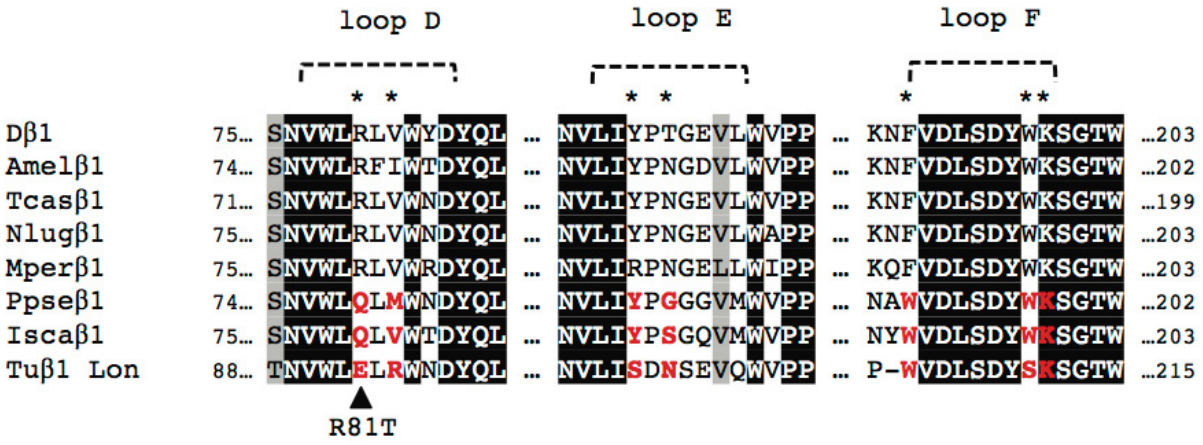


Figure VI.3 - Alignment of the amino acid sequences in loop D, E and F of nicotinic acetylcholine receptor β1 subunits of *D. melanogaster* (Dβ1), *A. mellifera* (Amelβ1), *T. castaneum* (Tcasβ1), *Nilaparvata lugens* (Nlβ1), *M. persicae* (Mperβ1), *P. pseudoannulata* (Ppseβ1), *I. scapularis* (Iscaβ1) and London strain of *T. urticae* (Tuβ1 Lon). Fully conserved identical residues are shaded in black, while similar residues are shaded in grey. An asterisk marks the residues (R81, V83, R135, N137, F190, W197, K198 according to *M. persicae* nAChR β1 numbering) involved in imidacloprid selectivity (Shimomura *et al.* 2006; Yao *et al.* 2008; Song *et al.* 2009), with those of Chelicerata indicated in red. A R81T substitution, strongly associated with neonicotinoid resistance in *M. persicae* (Bass *et al.* 2011), is indicated with a triangle. See Table VI.2 for accession numbers.

3.3 *T. urticae* chloride channel and uncharacterized cysLGIC subunits

We identified 19 cysLGICs subunits other than nAChRs in the *T. urticae* genome (Table VI.4), which is a remarkably higher number compared to that found in insect genomes sequenced to date (e.g. 13, 10, 12 and 10 in *D. melanogaster*, *A. mellifera, T. castaneum* and *N. vitripennis*, respectively (Jones and Sattelle 2006; Jones and Sattelle 2007; Jones *et al.* 2010; Knipple and Soderlund 2010)).

3.4 *T. urticae* GABA-gated chloride channel subunits

The GABA receptors mediate inhibitory chloride currents and are major components of synaptic inhibition in the central nervous system of invertebrates and vertebrates (Hosie *et al.* 1997). Three genes (*Rdl*, *Grd* and *Lcch3*) that encode GABA-gated ion channel subunits have been identified in *D. melanogaster* (Knipple and Soderlund 2010). Orthologues of these genes were also detected in other insects like *A. mellifera*, *T. castaneum* and *N. vitripennis* (Jones *et al.* 2010). While most insect genomes sequenced to date contain only one *Rdl* orthologue (Jones *et al.* 2010; Yu *et al.* 2010), three *Rdl* orthologues were identified in the *T. urticae* genome (Fig. VI.1). In contrast, no orthologues of *Grd* and *Lcch3* were found in *T.*

urticae. Alternative splicing in insect *Rdl* has been documented and might possibly contribute to the functional variation in these receptors in insects (Jones *et al.* 2010). However, no alternative splice sites and exons could be detected in *T. urticae Rdl* genes and it is possible that the number of *Rdl* orthologues itself might provide functional variability instead. In insects, all Rdl possess the PAR motif before transmembrane region 2 (TM2), which is important for forming anion channels, while Grd and Lcch3 lack this sequence, and may form cation permeable channels (Jensen *et al.* 2005; Jones *et al.* 2010). In *T. urticae* all Rdl orthologues possess the PAR-motif, indicating that only GABA-gated anion channels may be formed. While *Tu_Rdl1* is located on scaffold 12, *Tu_Rdl2* and *Tu_Rdl3* are tightly clustered (within 2.2 kb of each other) and oriented in opposite direction on scaffold 36, suggesting a duplication event. Interestingly, Gypsy LTR transposons (*tetur36te00033*, *tetur36te00034, tetur36te00035*, *tetur36te00036, tetur36te00037, tetur36te00038*) are located between exon 4 and exon 5 of *Tu_Rdl2* and a PiggyBac (*tetur36te00039*), LINE (*tetur36gte00040*) and Mutator transposon (*tetur36te00041)* were identified between exon 3 and 4, exon 7 and 8 and next to the 3'end of *Tu_Rdl3*, respectively. All three *T. urticae Rdl* orthologues are transcribed in all developmental stages, with *Tu_Rdl3* showing slightly higher expression levels compared to the other two homologs (Fig. VI.2).

GABA receptors are validated targets for the cyclodienes (such as dieldrin) and phenylpyrazole (such as fipronil) insecticides, which have been used extensively for pest control in agricultural and public health (De Jong 1991; Narahashi *et al.* 2010). A resistance substitution, (A301S) (and the alternative substitutions A301G, A301N) that confers resistance to dieldrin and fipronil, has since long been identified and functionally characterised in the protein encoded by *Rdl* (Resistant to dieldrin) in a wide range of insects (Buckingham *et al.* 2005; Hansen *et al.* 2005; Li *et al.* 2006; Nakao *et al.* 2010). We found that all three *T. urticae* Rdl orthologues in *T. urticae* strains from different geographic origin (London, MAR-AB, 033, LS-VL), have no alanine at the conserved position 301, but instead the resistance associated serine (Tu_Rdl2 and Tu_Rdl3) or a histidine (Tu_Rdl1). In a laboratory selected fipronil resistant *Drosophila* strain, another substitution (T350M) has been detected in combination with A301S. Functional expression revealed that T350M alone reduces the sensitivity to fipronil and dieldrin in Rdl, while the double mutant (A301S+T350M) showed the highest resistance to fipronil (Le Goff *et al.* 2005). A substitution at this conserved position in insects (T350A) was present in Tu_Rdl1 of all examined *T. urticae* strains. In addition, it was recently shown that a new substitution

(T305L) was associated with dieldrin resistance in the cattle tick *R. microplus* (Hope *et al*., 2010), and a similar residue (I305) was also detected in Tu_Rdl1 of all strains examined (Fig. VI.4).

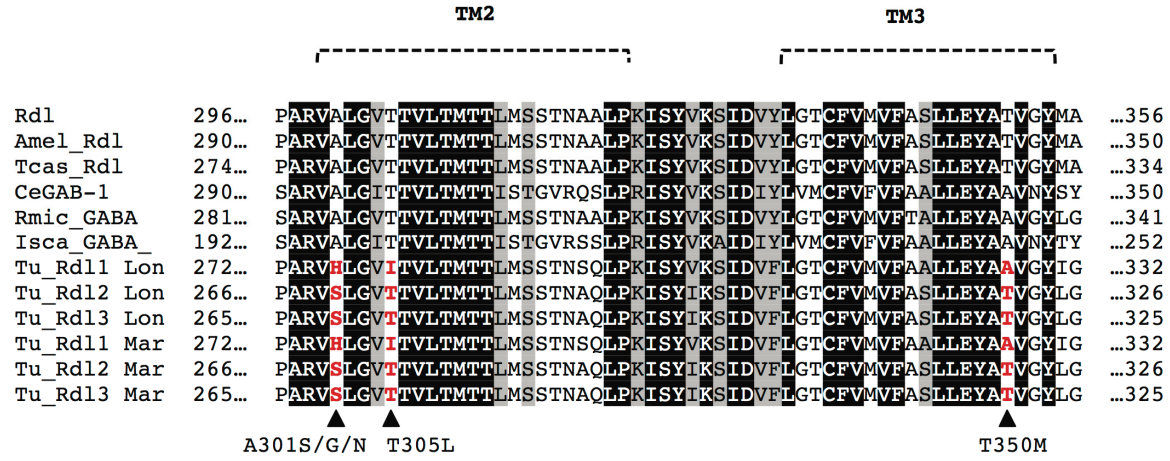


Figure VI.4 - Alignment of the amino acid sequences in transmembrane domain 1 and 2 (TM1 and 2, respectively) of γ-aminobutyric acid (GABA)-gated chloride channel subunits of *D. melanogaster* (Rdl), *A. mellifera* (Amel_Rdl), *T. castaneum* (Tcas_Rdl), *C. elegans* (CeGAB-1), *R. microplus* (Rmic_GABA), *I. scapularis* (Isca_GABA) and London (Lon) and MAR-AB (Mar) strain of *T. urticae* (Tu_Rdl1, Tu_Rdl2, Tu_Rdl3). Fully conserved identical residues are shaded in black, while similar residues are shaded in grey. The A301G/S/N, T305L, and T350M substitutions (*D. melanogaster* Rdl numbering*)*, associated with dieldrin and/or fipronil resistance in arthropods (Le Goff *et al.* 2005; Hope *et al.* 2010; Nakao *et al.* 2010) are marked with a triangle, while residues at equivalent positions in *T. urticae* are indicated in red. For accession numbers see Table VI.2.

All together, the above sequence features of *T. urticae* Rdls might explain the reduced sensitivity of spider mites to fipronil; mortality across all four strains used in this study was less than 10%, 24h post exposure at field dose of 200 mg/L (Fig. VI.5). In contrast to spider mites, ticks, which have the residue associated with susceptible phenotypes at position 301 (301A), are effectively controlled by fipronil (George *et al.* 2004). Also, in literature there are no recent reports describing that cyclodienes, such as endosulfan and dienochlor, can still effectively control spider mites (Gough 1990; Wilson *et al.* 1998). This might be due to the fixation of the A301S/H substitutions in many areas as indicated by our analysis in a number of strains from different regions (data not shown). It remains to be determined whether this is due to the extensive cyclodiene selection pressure in the past decades (particularly between 1950-1990) (De Jong 1991).

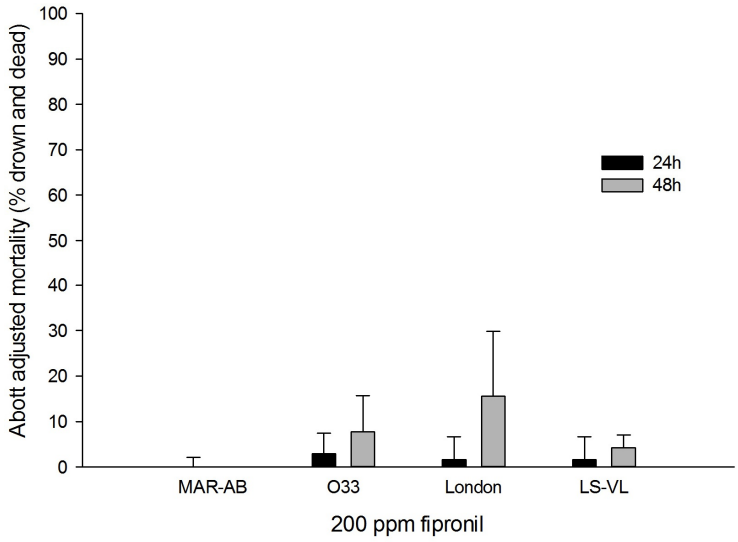


Figure VI.5 - Toxicity of 200 ppm fipronil to four different strains of *T. urticae*

3.5 *T. urticae* histamine-gated chloride channel subunits

Histamine-gated chloride channels are mainly expressed in arthropod eyes and are activated by histamine, which has been recognized for several years as the major neurotransmitter of arthropod photoreceptors (Zheng *et al.* 2002). In *D. melanogaster*, *A. mellifera, T. castaneum* and *N. vitripennis*, two HisCl genes are present (Jones and Sattelle 2006; Jones and Sattelle 2007; Jones *et al.* 2010; Knipple and Soderlund 2010). In contrast, four putative HisCl genes (*Tu_HisCl1*, *Tu_HisCl2*, *Tu_HisCl3* and *Tu_HisCl4*) were identified in the *T. urticae* genome (Fig. VI.1, Appendix VI-B). All four *T. urticae* HisCls are transcribed in at least one life stage, although the expression of *Tu_HisCl4* is very low (Fig. VI.2). The PAR-motif or a variation thereof (SAR in Tu_HisCl1) preceding TM2 was detected in all HisCl, which is consistent with their function as anion channels (Jensen *et al.* 2005; Jones *et al.* 2010). There is one clear *T. urticae* orthologue (Tu_HisCl1) of the insect HisCl subunit, while the other three *T. urticae* HisCls (Tu_HisCl2, Tu_HisCl3 and Tu_HisCl4) form a sister clade to insect HisCl subunits and Tu_HisCl1 (Fig. VI.1). As most chelicerates have only ocelli and insects have sophisticated compound eyes (whether or not combined with ocelli) (Oakley 2003), this expansion of HisCl genes in *T. urticae* is intuitively difficult to explain. However, HisCl subunits might also be involved in regulation of temperature preference and cold tolerance (as shown for *D. melanogaster* (Hong *et al.* 2006)) and would thus not be restricted to the visual system of *T. urticae*. Some studies indicated a potential role of HisCl in the neurotoxic action of avermectins, although this remains to be further investigated (Georgiev *et al.* 2002; Iovchev *et al.* 2002).

3.6 *T. urticae* pH-sensitive chloride channel subunits

The pH-sensitive chloride channel subunit was first identified in *Drosophila* (Schnizler *et al.* 2005) and subsequently, orthologues of this cysLGIC subunit were detected in every insect genome sequenced to-date (Jones *et al.* 2010). We identified a single clear pHCl orthologue with a characteristic PAR-sequence preceding TM2, that is transcribed in all developmental stages (Fig. VI.2, Appendix VI-B). Several studies (Schnizler *et al.* 2005; Mounsey *et al.* 2007) showed that pHCl is activated in a pH dependent way by ivermectins.

3.7 Uncharacterized cysLGIC subunits

No orthologues of Insect group I cysLGIC genes were found in *T. urticae*, which suggests that these cysLGIC genes evolved after the split of insects and chelicerates and are indeed limited to insects as denoted by Jones *et al*. (2010). In contrast, *T. urticae* possesses four cysLGIC subunits (Tu_12344.1, Tu_12344.2, Tu_12344.3 and Tu_12344.4) which form a sister clade with uncharacterized cysLGIC subunits of *D. melanogaster*, *T. castaneum* and *A. mellifera*, with high bootstrap support (93%) (Fig. VI.1). Remarkably, all members of this clade lack the PAR amino acid motif, suggesting that they might not form anion channels (Jensen *et al.* 2005). Together with Tu_Rdl1, Tu_Rdl2, Tu_Rdl3, these cysLGIC subunits also lack the two additional cysteines in Loop C (Appendix VI-B), a common feature in all glutamate, histamine and pH-sensitive gated chloride channels and vertebrate glycine gated receptor subunits (Appendix VI-B, (Dent 2006)). Due to a frameshift, only a partial cysLGIC gene could be annotated for *Tu_12344.2*, with loop A missing, although this gene is the highest transcribed of all uncharacterized cysLGIC subunits (Fig. VI.2). Finally, one *T. urticae* gene (*Tu_LGIC_unk*) with cysLGIC characteristics (Appendix VI-B) is expressed throughout the lifecycle (Fig. VI.2) but has no arthropod orthologues (Fig. VI.1). It remains to be determined if any of these uncharacterized cysLGIC subunits play a role in insecticide/acaricide toxicology.

3.8 *T. urticae* glutamate-gated chloride channel subunits

Most insects such as *D. melanogaster*, *T. castaneum* and *A. mellifera*, have only one glutamate-gated chloride channel subunit. Surprisingly, we identified six orthologous GluCl genes (*Tu_GluCl1*, *Tu_GluCl2*, *Tu_GluCl3*, *Tu_GluCl4*, *Tu_GluCl5* and *Tu_GluCl6*) in the genome of *T. urticae* (Table VI.4, Fig. VI.1). Consistent with their function as anion channel, all *T. urticae* GluCl subunits have the PAR motif or a variation thereof (AAR in Tu_GluCl1) (Jensen *et al.* 2005) and all display high similarity with the *D. melanogaster* orthologue

(68.4%-76.1%) (Appendix VI-C). Two GluCl genes (*Tu_GluCl6* and *Tu_GluCl5*) are 99.2% identical and surrounded by an identical stretch of genes on different scaffolds, indicating a recent duplication event. The possibility that *Tu_GluCl6* and *Tu_GluCl5* represent the same gene, due to a mis-assembly caused by allelic variants, cannot be excluded at this stage. With the exception of *Tu_GluCl5* and *Tu_GluCl6*, all *T. urticae* GluCl genes are transcribed at relatively high levels throughout the spider mite lifecycle (Fig. VI.2).

The high number of GluCl genes is in line with a previous study in the nematode *C. elegans* (Jones and Sattelle 2008). *T. urticae* GluCl subunits display a high amino acid variability at positions that also vary between alternative spliced exons of insect GluCl subunits (e.g. between exon 3a and 3b of GluCl of *D. melanogaster*) (Appendix VI-B). Since alternative splicing was not detected in the *T. urticae* GluCl subunits, it is possible that the expansion of those receptors contributes to the functional diversity possibly required in *T. urticae*.

Glutamate chloride channels are targets of macrocyclic lactone natural products (i.e. avermectins and milbemycins) in both insects and nematodes (Cully *et al.* 1994; Kane *et al.* 2000; Wolstenholme 2010). This chemical class has been largely employed in pest control, as well as animal and human health applications, with worldwide annual sales of approximately US $2 billion sales (Wolstenholme 2010). Binding of avermectin molecules to GluCls causes irreversible opening of the channels, leading to an influx of chloride ions, hyperpolarisation and paralysis (Mounsey *et al.* 2007). Furthermore, GluCls may also play a secondary role in mediating fipronil toxicity, as this has been documented in German cockroaches (Narahashi *et al.* 2010). Several studies on GluCls have identified different substitutions that affect ivermectin sensitivity. In *D. melanogaster* the P299S substitution of GluCl has been associated with low levels of ivermectin resistance (Kane *et al.* 2000), while three (E114G, V235A and L256F) and two (V60A and R101H) abamectin resistance substitutions have been reported in the GluCla3 and GluClb subunit of the nematode *Cooperia onchophora*, respectively (Njue *et al.* 2004). A substitution in the highly conserved TM2 region of Tu_GluCl1 (G323D) was strongly associated with a 17.9-fold reduction in abamectin sensitivity in an isogenic resistant strain of *T. urticae* (Kwon *et al.* 2010). A substitution at this position was functionally characterized in a vertebrate glycine receptor, and shown to cause a dramatic reduction in avermectin sensitivity (Lynagh and Lynch 2010).

3.9 Identification of a novel mutation in *Tu_GluCl3* associated with abamectin resistance in *T. urticae*

Given the previously unknown expansion of GluCl subunits revealed in this study, we investigated the possible presence of alternative and/or additional resistance mutations in a highly resistant *T. urticae* strain (MAR-AB) to abamectin (resistance levels >2,000-fold, Table VI.6). Inhibitors of known detoxifying enzymes only partially synergized resistance (data not shown), suggesting that target-site resistance could be a main resistance mechanism. Reciprocal crosses between the abamectin susceptible London strain and MAR-AB, revealed that resistance was incompletely recessive and not maternally inherited (Table VI.6). F1 female mites of reciprocal crosses were back-crossed and the dose-response relationship of abamectin toxicity in F2 female mites revealed that resistance was not monogenic (for F2b (F1b ♀ x S ♂) χ^2= 14.63 ($p < 0.01$, df=4); and for F2a (F1a ♀ x R ♂) χ^2= 38.14 ($p < 0.001$, df=4) (Table VI.6)). Hence, the inheritance of abamectin resistance in *T. urticae* may be controlled by more than one incompletely recessive factor, which confirms earlier findings on abamectin resistance genetics in the closely related spider mite *Tetranychus cinnabarinus* (He *et al.* 2009).

Table VI.6 - Concentration probit-mortality data of abamectin on London (S) and MAR-AB (R) strain and estimation of degree of dominance (DD) in the female progenies.

Strain	n[a]	LC_{50} (ppm)	95% CI[b]	Slope (±SE)	χ^2 (df)[c]	RR[d]	DD[e]
London (S)	600	0.23	0.20 - 0.24	3.7 (±0.4)	10.5 (18)	1	
MAR-AB (R)	543	475	394 - 518	6.1 (±1.3)	19.1 (13)	2,065	
F1a (RR ♀ x S ♂)	526	9.1	7.1 - 11.6	1.4 (±0.1)	11.1 (13)	39.6	-0.03
F1b (SS ♀ x R ♂)	528	8.2	6.4 - 10.3	1.6 (±0.14)	11.0 (13)	35.6	-0.05
F2a (F1a ♀ x R ♂)	426	30.8	23.0 – 42.0	1.8 (±0.2)	13.4 (10)	133.9	
F2b (F1b ♀ x S ♂)	513	1.6	1.2 – 2.0	1.5 (±0.1)	3.9 (13)	7.0	

[a] numbers of mites used
[b] 95% Confidence Interval
[c] χ^2 value and degrees of freedom (df) as calculated by PoloPc
[d] RR= Resistance Ratio
[e] Degree of Dominance calculated according to Stone 1968

The observed polygenic nature of target site abamectin resistance is particularly interesting in the light of the multiple GluCl and Rdl genes described above, and consequently we re-sequenced these genes in the MAR-AB strain (GenBank accession numbers: JQ738191-JQ738198). A multiple alignment of all GluCl subunits of London and MAR-AB revealed

that in both strains a glutamate was present in Tu_GluCl4 and Tu_GluCl5/6 at the same site as reported by Kwon *et al*. (2010).

However, only in the MAR-AB strain we found the previously described substitution G323D (equivalent with G314 in our annotation of Tu_GluCl1) in Tu_GluCl1. In addition, we identified a novel substitution at an identical position (G326E) in Tu_GluCl3 of the MAR-AB strain (Fig. VI.6). In all other positions known to be involved in abamectin susceptibility, including E114 V235 L256 V60 and R101 (*Cooperia* numbering, (Njue *et al.* 2004)) and P299S (*Drosophila* numbering, (Kane *et al.* 2000)), we found identical residues in both the susceptible and resistant strain. No amino acid differences were detected between homologous Rdl receptors of the resistant and the susceptible strains. Hence, our findings confirm the previously identified substitution in Tu_GluCl1, but also indicate that an additional substitution in Tu_GluCl3 might have a similar or additive effect, as MAR-AB has a more striking resistance phenotype when compared to the abamectin resistant strain of Kwon *et al*. (2010). To investigate the relative role of these substitutions in abamectin resistance, we analysed the Tu_GluCl1 G314D and Tu_GluCl3 G326E genotype frequencies in individual mites of backcrossed F2 females (Tables VI.6 and VI.7). First, a χ^2square genetic linkage test on unselected F2a and F2b progenies (see Table VI.7 for details on crosses and names) revealed that G314D in Tu_GluCl1 and G326E in Tu_GluCl3 are independently inherited (Table VI.7). Next, survivors of the F2a progeny (genotypes RS and RR, LC_{50}=30.8 mg/L, Table VI.6) selected with 100 mg/L abamectin clearly showed a high enrichment of the homozygous Glu/Glu Tu_GluCl3 genotype (93.3 %), indicating that this substitution is recessive and strongly linked to abamectin resistance at high doses (Table VI.8).

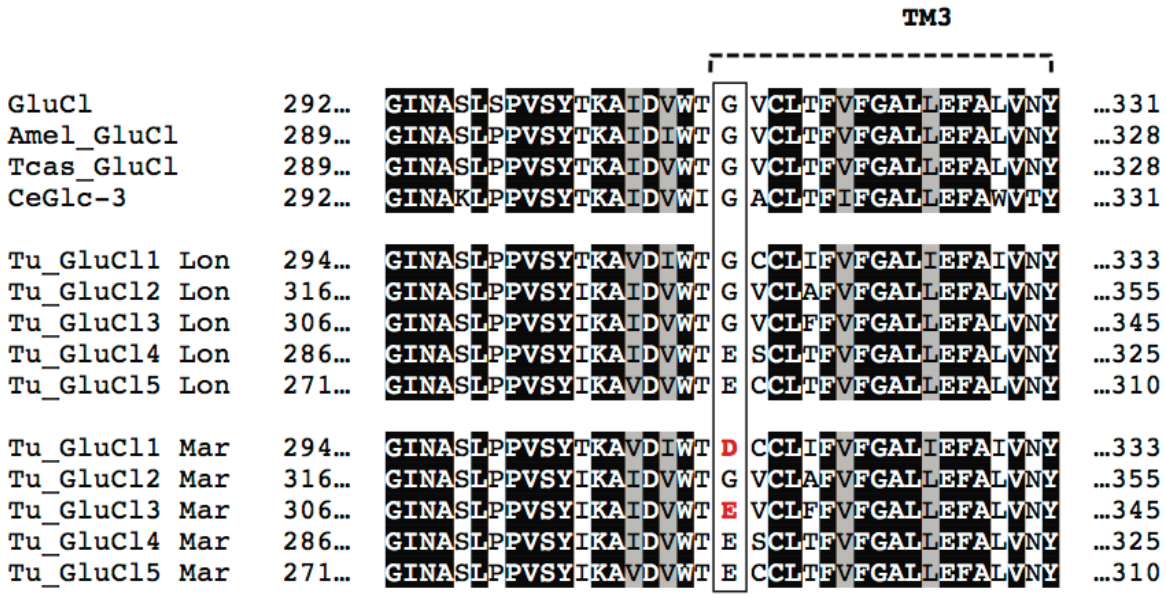


Figure VI.6 - Alignment of the amino acid sequences in transmembrane domain 3 (TM3) of glutamate-gated chloride channel subunits of *D. melanogaster* (GluCl), *A. mellifera* (Amel_GluCl), *T. castaneum* (Tcas_GluCl), *C. elegans* (CeGlc-3) and London (Lon) and MAR-AB (Mar) strain of *T. urticae* (Tu_GluCl1, Tu_GluCl2, Tu_GluCl3, Tu_GluCl4, Tu_GluCl5).

Fully conserved residues are shaded in black, while similar residues are shaded in grey. The G314D and G326E substitutions in Tu_GluCl1 and Tu_GluCl3, respectively, of the *T. urticae* MAR-AB strain, associated with abamectin resistance, are indicated in red ((Kwon *et al.* 2010), this study). For accession numbers see Table VI.2

Table VI.7 - χ^2 goodness of fit test for independent inheritance of G314D and G326E substitutions in Tu_GluCl1 and Tu_GluCl3, respectively

	n[a]	Tu_GluCl1/Tu_GluCl3[b]				$\chi^{2\,c}$
		AspAsp/GluGlu	GlyAsp/GluGlu	GlyAsp/GlyGlu	AspAsp/GlyGlu	
(F1a ♀ x R ♂)[d] unselected	20	25.0	30.0	25.0	20.0	0.4
		GlyGly/GlyGlu	GlyAsp/GlyGly	GlyAsp/GlyGlu	GlyGly/GlyGly	
(F1b ♀ x S ♂)[d] unselected	20	15.0	40.0	25.0	20.0	2.8

[a] number of individuals used for genotyping
[b] percentage of genotypes observed in Tu_GluCl1 and Tu_GluCl3
[c] χ^2critical= 7.82, df=3, p=0.05
[d] R= abamectin resistant strain (MAR-AB); S= abamectin-susceptible strain (London); F1a = (RR ♀ x S ♂); F1b = (SS ♀ x R ♂), see Table VI.6

In contrast, the Gly/Asp and Asp/Asp genotypes for Tu_GluCl1 were found in a 50/50 relationship, and thus G314D in Tu_GluCl1 is dominant or is redundant for survival at high abamectin doses. Survivors of the F2b cross (genotypes SS and RS, LC_{50} of 1.6 mg/L, Table VI.6) selected with a lower abamectin dose (10 mg/L) showed a frequency of 75% for the heterozygous Gly/Asp Tu_GluCl1 genotype and 87.5% for the heterozygous Gly/Glu Tu_GluCl3 genotype (Table VI.8). Thus, at lower abamectin concentration, both substitutions are associated with resistance. Mites homozygous for Gly in both Tu_GluCl1 and Tu_GluCl3 were never detected amongst the survivors of either dose tested. Given the fact that both *Tu_GluCl1* and *Tu_GluCl3* are highly expressed (Fig. VI.2) and substitutions in only these subunits were detected and linked with resistance, Tu_GluCl1 and Tu_GluCl3 might be the prime target-sites of abamectin in *T. urticae*.

Table VI.8 - Genotyping of the G314D and G326E substitutions in F2 progenies unselected and selected with 100 ppm or 10 ppm abamectin.

	n[a]	Tu_GluCl1		(%)[b]	χ^2 [c]	Tu_GluCl3		(%)[d]	χ^2 [c]
		Gly/Asp	Asp/Asp			Gly/Glu	Glu/Glu		
(F1a ♀ x R ♂)[e] 100 ppm selected	30	50.0	50.0	75.0	0.0	6.7	93.3	96.7	22.6
(F1a ♀ x R ♂)[e] unselected	20	55.0	45.0	72.5	0.2	45.0	55.0	77.5	0.1
		Gly/Asp	Gly/Gly			Gly/Glu	Gly/Gly		
(F1b ♀ x S ♂)[e] 10 ppm selected	24	75.0	25.0	37.5	6.0	87.5	12.5	43.7	13.5
(F1b ♀ x S ♂)[e] unselected	20	65.0	35.0	32.5	1.8	40.0	60.0	20.0	0.8

[a] number of individuals used for genotyping
[b] G314D frequency
[c] χ^2 critical = 3.84, df=1, p=0.05
[d] G326E frequency
[e] R= abamectin resistant strain (MAR-AB); S= abamectin-susceptible strain (London); F1a = (RR ♀ x S ♂); F1b = (SS ♀ x R ♂), see Table VI.6

4 Conclusions

We described and analyzed the cysLGIC gene family of *T. urticae*, the largest one among arthropods up to date and the first described family within the subphylum of the Chelicerata. Phylogenetic analysis and sequence comparison with insects revealed that spider mites have a surprisingly high number of *Rdl* and GluCl genes, but similar numbers of other cysLGIC subunits, such as the nAchRs. By analysing sequences of known insecticide targets, we found remarkable similarities between insensitive (resistant) insect cysLGIC forms and *T. urticae* orthologues, which might be associated with the low toxicity of certain insecticides (such as neonicotinoids and fipronil) against spider mites. We further showed that on top of a previously documented substitution, a novel G326E substitution in Tu_GluCl3 is strongly associated with high levels of abamectin resistance.

Chapter VII

Conclusions and Future Perspectives

The subclass of the Acari, mites and ticks, has always been less studied than the Insecta. This bias is also observed in the genomics era of arthropod research (Chapter I). This study aims at delivering a small contribution to adjust this discrepancy. This was accomplished, firstly, by elucidating the mitochondrial (mt) genome sequences of two economically important mites (Chapter II and III) and secondly by taking full advantage of the deciphered genome sequence of the two spotted spider mite *Tetranychus urticae*. This genome sequence was the starting point for the design of a whole genome gene expression microarray which was used to study xenobiotic adaptation in spider mites (Chapter IV) but also led to the characterization of *T. urticae* gene families either known to be involved in detoxification (Chapter V) or as possible target-sites of pesticides (Chapter VI).

Mite mitochondrial genomes

Using a PCR amplification strategy the complete mt genome of an acariform species, the house dust mite *Dermatophagoides pteronyssinus* (Chapter II) and a mesostigmatid species, the predatory mite *Phytoseiulus persimilis* (Chapter III) was obtained. A restriction digest of rolling circle amplified mt DNA confirmed the mt genome size of both mite species. This technique for assessing mt genome length is not frequently used by researchers working on mt genomes. Nevertheless, it could be a helpful tool in detecting non-specific amplification artefacts, which are well-known drawbacks of a PCR-approach. Also, in the light of many ongoing genome projects using next generation sequencing techniques, the mitochondrial genome can be assembled from the sequenced short DNA reads. However, depending on the heterogeneity of the starting material used, this might lead to artefacts in assembly, which can be ruled out with rolling circle amplification and a restriction digest.

Mitochondrial genome length, gene and AT-content of *D. pteronyssinus* and *P. persimilis* were similar to previously sequenced mite mt genomes. The mt gene orders of both *D. pteronyssinus* and *P. persimilis* were highly rearranged compared to the ancestral arthropod pattern and represented new mt gene patterns within the Arthropoda. A comparison of Acari mt gene orders indicated that this genome "character" seems less useful for deduction of phylogenetic relationships between Acari superorders, confirming an earlier hypothesis of Fahrein *et al*. (2007). Next to *D. pteronyssinus* and *P. persimilis* many other, mainly acariform, mite species exhibit highly shuffled mt gene orders (Yuan *et al.* 2010, Edwards *et al.* 2011, Chapter II). This is in contrast to most arthropod species, having a highly conserved mt gene pattern (Stach *et al.* 2010). However, these mite mt genomes with extreme gene

rearrangements belong to species distributed over only five acariform families and one mesostigmatid family. A broader taxonomic sampling of mite mt genomes should be performed to reveal whether the occurrence of such highly rearranged mt gene orders is widespread within the Acari. With the current sequencing technologies (e.g. NGS technologies, sequencing of minute amounts of DNA) such sampling should be affordable. Next to extreme mt gene arrangements, the strand bias in nucleotide composition (measured by GC- and AT-skews) from the mt genomes of both *D. pteronyssinus* and *P. persimilis* was reversed compared to those found in most metazoan mt DNA. It has been suggested that a strand swap of the control region, containing all initiation sites for transcription and replication of mtDNA, could result in the reversal of GC-and AT-skews (Hassanin *et al.* 2005). As there is still much debate about the mode of mtDNA replication (Bernt *et al.* 2012), this hypothesis is speculative and warrants further investigation.

In the case of the *D. pteronyssinus* mt genome, extremely truncated mt tRNAs were identified (Chapter II). These tRNAs lack either a D- or T-arm and hence deviate from the typical cloverleaf tRNA structure. Similar tRNA structures were also identified in the closely related *D. farinae* and other acariform species (Yuan *et al.* 2010, Klimov *et al.* 2010). Intriguingly, *D. pteronyssinus* and other acariform rRNAs were also shortened compared to those of parasitiform mites and most other metazoan species (Chapter II, Yuan *et al*. 2010, Klimov *et al*. 2011). Wolstenholme *et al.* (1987) suggested that truncated mt rRNAs and atypical mt tRNAs might have evolved in a concerted fashion in nematodes. Whether this hypothesis is true, and, by extension, also applies for acariform species, has been poorly investigated (Ohtsuki *et al*. 2007) and needs further study. Furthermore, we also detected a microsatellite in the mt genome of *D. pteronyssinus*. This is remarkable since a mt microsatellite was never reported before in species belonging to the Chelicerata. In the absence of a complete *D. pteronyssinus* genome, this genetic marker could be useful in population genetic studies of house dust mites.

In contrast to *P. persimilis*, the mt genome of the closely related *Metaseiulus occidentalis* was reported to be extremely large, to lack *nad6* and *nad3* protein coding genes and to contain 22 tRNAs without T-arms (Chapter III). Puzzled by these differences, additional experiments were performed on the mt genome of *M. occidentalis*. It was found to be smaller than previously reported, it does not lack a *nad3* gene and at least 15 out of 22 tRNAs can be folded into canonical cloverleaf structures. The discrepancies between the reported mt

genome in literature and our data might be caused by the use of different lines of this species in experiments. However, it seems more likely that a misassembly of contigs was the main cause, and in this case, the rolling circle amplification technique has allowed a more accurate genome size estimation. This also illustrates how mt genome comparison of closely related species can help in the fine-tuning of mt genome annotation. Hence, new mt genome sequencing projects should preferably not only sequence the mt genome of the species of interest but also sequence the mt genome of their closest relatives. This approach would certainly be affordable, given the current low cost sequencing technologies.

Finally, a maximum likelihood phylogenetic analysis using concatenated Acari mt gene sequences resulted in recovering Acari order relationships concordant to traditional views of Acari phylogeny. The monophyletic nature of the subclass Acari however was not investigated in this study. This feature has been questioned for more than 60 years (Dunlop and Alberti 2008). Recent studies, using a phylogenetic analysis of chelicerate mt genome data, failed in recovering the monophyly of the Acari (Domes *et al.* 2008, Ovchinikov and Masta 2012, Rota-Stabelli *et al*. 2013), probably due to the fact that only few mite taxa were included in these studies. A thorough mt genomic analysis focusing on Acari and their closest relatives, with sampling in a wider group of taxa, could help in finding a definitive answer in this long-debated phylogenetic issue. On the other hand, the phylogenetic signal of mt genomes is maybe too limited for resolving arthropod interordinal relationships, as suggested by Talavera *et al*. (2011), and hence (slower evolving) nuclear genes/genomes could be more appropriate for resolving the monophyly of the Acari.

The spider mite genome

Gene expression microarray

The two-spotted spider mite *T. urticae* is an extreme polyphagous pest with more than 1,100 documented host plants. It also has an extraordinary ability to develop pesticide resistance. The availability of the complete genome sequence now (Grbić *et al.* 2011) provides unique information and tools to study the molecular mechanisms behind these phenomena. A whole-genome expression microarray was developed based on the *T. urticae* genome sequence and used to follow changes in *T. urticae* gene expression after host transfer and in pesticide-susceptible versus pesticide-resistant strains (Chapter IV). When mites from a pesticide susceptible strain propagated on bean were transferred to a more challenging host (tomato), transcriptional responses increased over time with about 7.5 % of genes differentially expressed after five generations. Whereas many genes with altered expression belonged to known detoxification families (like P450 mono-oxygenases (P450s)), new gene families previously not associated with detoxification in other herbivores showed a striking response, including the lipocalin family, the major facilitator superfamily (MFS) and intradiol ring-cleavage dioxygenases (ID-RCDs). Lipocalins are involved in binding small molecules and are thought to mediate interactions with plant compounds. The MFS genes encode transporters that might remove toxins or their metabolites from cells. Phylogenetic analysis showed that ID-RCDs were recently acquired from either bacteria or fungi via horizontal gene transfer. Their proliferation in the *T. urticae* genome, and strong transcriptional response to host transfer, is consistent with selection for an expanded "toolkit" for the metabolism of foreign compounds by these spider mites, even though the biochemical targets of these enzymes are currently unknown. Consistent with the changes in gene expression that were observed after host transfer, a large number of genes differed in expression between pesticide-susceptible and pesticide-resistant spider mites. Strikingly, these differences also occurred in many of the same genes, or genes belonging to the same family, affected by host transfer, such as genes encoding P450s, lipocalins, MFS, ID-RCDs and many genes of unknown function that are predicted to encode for secreted proteins. In particular, the direction of change in expression was often correlated for given genes. Obtaining functional evidence that members of the above mentioned families (ID-RCDs, lipocalins, MFS, secreted proteins) actually contribute to xenobiotic tolerance when up-or downregulated should be the priority of further research. Furthermore, since the genome sequences of the closely related but less polyphagous spider mite *Tetranychus evansi* and the monophagous gorse spider mite

Tetranychus lintearius will soon become available (Rombauts *et al*. 2013), it would also be interesting to investigate how these gene families evolved in a monophagous, oligophagous and polyphagous spider mite species.

Because the transcriptional patterns observed for host transfer and pesticide resistance were globally similar with regards to both gene type and direction of expression changes, it was tested whether adaptation of susceptible mites to tomato was sufficient to increase tolerance to pesticides. Adaptation to tomato indeed decreased the toxicity of three out of five acaricides tested. However, it was not yet tested whether the development of resistance can influence the ability of mites to cope with new host plants and this should be a priority in further documenting a parallel between host-plant range and pesticide resistance. Also, although mites grown for five generations on tomato were considered adapted to tomato, classical validation of adaptation was not experimentally tested in this study. This can be easily investigated by transferring mites feeding on tomato to bean for a few generations, and subsequently compare their fitness on tomato in comparison with the parental line. The effect of adaptation to tomato on the fitness on other host plants might also be interesting to investigate. It could potentially point to mechanisms of adaptation, as it is not known whether adaptation to a single host erodes the genetic variability, and whether this variability is crucial in plant adaptation. Also, our study compared gene expression from mites feeding on beans, to that of mites feeding on tomato for 5 generations, and did not allow to determine which differences in gene expression are induced and which are constitutive. In this context, it might be interesting to investigate whether adaptation to tomato influences the inducibility of genes. Finally, it is also not clear what the underlying causes of the observed gene expression differences are. Do pronounced differences in gene expression reflect selection on standing genetic variation, or are alternatively physiological or epigenetic changes (like e.g. DNA methylation) at play?

The observed similarity in transcriptional responses between resistance and plant adaptation are interesting in the light of classical theory regarding the development of arthropod resistance to pesticides. Theory predicts that monogenic resistance under field conditions is caused by a rare allele that is typically selected far outside the range of susceptible phenotypes, and this has been well documented in field settings (Van Leeuwen *et al*. 2010a, McKenzie and Batterham 2004). However, theoretical studies on the evolution of pesticide resistance also have postulated that polygenic resistance (i.e. selection on many loci) can be expected when the range of the selective dose falls within the viability distribution of

susceptible phenotypes (McKenzie and Batterham 2004). The results of this study support a model in which adaptation to different hosts by polyphagous herbivores can generate various distributions of susceptible phenotypes for pesticides, thus facilitating survival in the face of pesticide application and hence the evolution of high resistance levels by either selection on standing variation or via the acquisition of mutations.

The ABC protein family

The ABC protein family is a well-known family involved in detoxification of xenobiotics, but has often been neglected in literature when a sequenced genome was used to investigate detoxification mechanisms (Claudianos *et al*. 2006, Oakeshott *et al*. 2010, Ramsey *et al.* 2010) Most members of this family either directly transport toxicants out of the cell or after conjugation with glutathione (Chapter V). Mining the genome of *T. urticae* for this family revealed the presence of 103 ABC genes. This is the highest number of ABC genes reported for a metazoan species to date. This high number is mainly due to lineage specific expansions of ABC genes within the ABCC, ABCG and ABCH subfamilies. Interestingly, most of the differentially expressed ABC genes in acaricide resistant strains and after introducing mites to a more challenging host belonged to these expanded ABC-subfamilies. This hints at their potential role in detoxification and may explain their retention after duplication in the mite genome. For future research, it would be interesting to validate the role of ABC proteins in spider mite detoxification pathways and obtain an inventory of the number of substrates/pesticides they transport. In the past *in vitro* assays using Sf9 cell lines have proven to be successful for this purpose (Özvegy *et al.* 2001). Besides these *in vitro* assays, RNAi could be another option to study ABC function in spider mites as was recently shown for the flour beetle *T. castaneum* (Broehan *et al*. 2013). Finally, similar as for the new gene families previously not associated with detoxification (see above), it would also be interesting to study the proliferation of ABC genes in the less polyphagous *T. evansi* and monophagous *T. lintearius*.

The cys-loop ligand gated ion channel family

Adaptation to pesticides can either occur by a mechanism of decreased exposure (e.g. by detoxification enzymes or ABC-transporters) or by a decreased response for these compounds e.g. by modifications in the target-site that alters pesticide binding. Members of

the cys-loop ligand gated ion channel family (cysLGIC) are well known as targets of pesticides and were characterized for the spider mite *T. urticae* (Chapter VI). The cysLGIC gene family of *T. urticae* represents the largest arthropod cysLGIC family to date. Phylogenetic analysis and comparison of the cysLGIC subunits with their counterparts in insects revealed that the *T. urticae* genome encodes for a high number of glutamate gated chloride channel (GluCl), histamine gated chloride channel (HisCl) and *Rdl* (*R*esistance to dieldrin) genes compared to insects but for a similar number of other cysLGIC subunits, like nicotinic acetyl choline receptors (nAChRs). Alternative splicing in insect cysLGIC genes has been well documented and might possibly contribute to the functional variation in these receptors in insects (Jones *et al*. 2007). Contrary to insects, no alternative splice sites could be detected in *T. urticae* cysLGIC genes, suggesting that in many cases the number of cysLGIC genes itself might provide functional variability instead. However, the experimental approach in this study was not specifically designed at detecting alternatively spliced genes with amplified cysLGIC fragments either directly sequenced or ligated into a cloning vector with 3-5 clones sequenced. In order to effectively identify alternatively spliced variants, recently available strand specific RNA sequencing data can be mined, which can provide more power to detect alternatively spliced reads and rare transcripts, as compared to PCR amplification and classical sequencing of clones. Next to alternative splicing, pre mRNA A-to-I editing also adds to functional diversity of cysLGICs in insects. As pre mRNA editing is widespread among animal tissues (Buckingham *et al.* 2005) and the gene coding for the main enzyme of this process is present in the spider mite genome (double-stranded RNA adenosine deaminase: *tetur09g03000* and *tetur01g10860*), it is most likely that pre mRNA editing also occurs in spider mites. However, instead of focusing on cysLGIC genes solely, the availability of the spider mite genome in combination with next generation sequencing technologies would also allow to study pre mRNA editing on a genome wide scale. As it is challenging to distinguish RNA editing events from genomic SNPs (Bass *et al.* 2012), the use of highly inbred spider mite strains will be a major advantage in identifying RNA-edited sites.

By analyzing sequences of known insecticide targets, we found remarkable similarities between insensitive (resistant) forms of nicotinic acetylcholine receptors and Rdl in insects and their orthologues in susceptible spider mite strains, which might be associated with the low toxicity of neonicotinoids and fipronil in spider mites. Functional expression in an *in vitro* assay system of these *T. urticae* orthologues before and after introduction of insect sensitive substitutions could demonstrate/validate the possible role of these substitutions in

the toxicity of these insecticides against spider mites (e.g. Song *et al*. 2009). Furthermore, we showed that on top of a previously documented substitution in Tu_GluCl1, a novel G326E substitution in Tu_GluCl3 is strongly associated with high levels of abamectin resistance in *T. urticae*. Genotyping these substitutions in abamectin resistant strains of the closely related *Panonychus citri* could further strengthen this association. Alternatively, introducing the mutations in a susceptible genetic background by repeated backcrossing (6-7x) might be a powerful approach to study the effect of a single mutation on the resistant phenotype without the confounding effects of additional resistance factors. A more direct approach to assess the effect of Tu_GluCL1 and Tu_CluCl3 substitutions in abamectin resistance is to functionally express channels in a model system (such as *Xenopus* oocytes) and determine the effect on channel currents *in vitro* (Cully *et al.* 1994). However, even without confirmation of their functional role in resistance, diagnostic tests could be developed to monitor the occurrence and spread of abamectin resistance in the field (Black IV and Vontas 2007).

As positively correlated cross-resistance between milbemectin and abamectin has been demonstrated, the role of these abamectin resistance associated substitutions in milbemectin resistance would also merit further investigation (Sato *et al.* 2005; Nicastro *et al.* 2010). Noteworthy, it was also shown that abamectin and milbemectin resistance in a laboratory selected and reared spider mite strain decreased in time in the absence of selection pressure. This instability of resistance is probably related to the occurrence of fitness costs associated with milbemectin/abamectin resistance in spider mites (Sato *et al.* 2005; Nicastro *et al.* 2010; Nicastro *et al.* 2011). Interestingly, such instability of resistance is considered favorable for the management of resistance (Dennehy *et al.* 1990). However, as other reports showed that abamectin resistance was stable for at least six month in a laboratory strain without selection pressure (Stumpf and Nauen 2002) instability of abamectin resistance cannot be considered as a generality, and might be related with the resistance mechanism present in a particular strain. Future studies should point out whether instability of abamectin resistance also occurs in the resistant strain under investigation in this study, and whether this instability is linked with the frequency of the detected mutations.

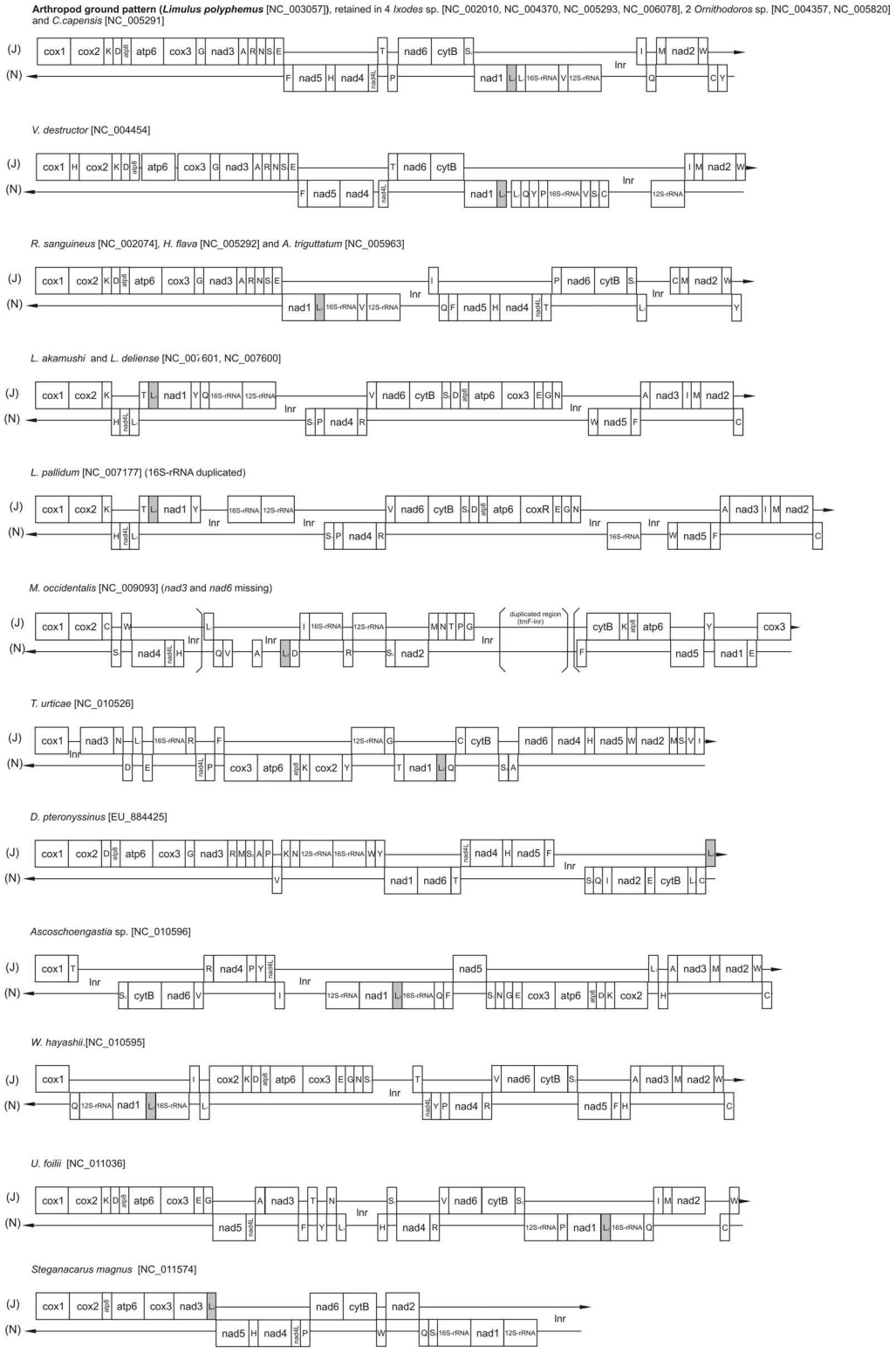
In conclusion…

This study delivers a small contribution to, Acari (post-)genomics. This new field of Acari research is however still in its initial phase and more Acari tailored protocols and methods (transformation of mites/ticks, efficient gene silencing in mites, long-term preservation of

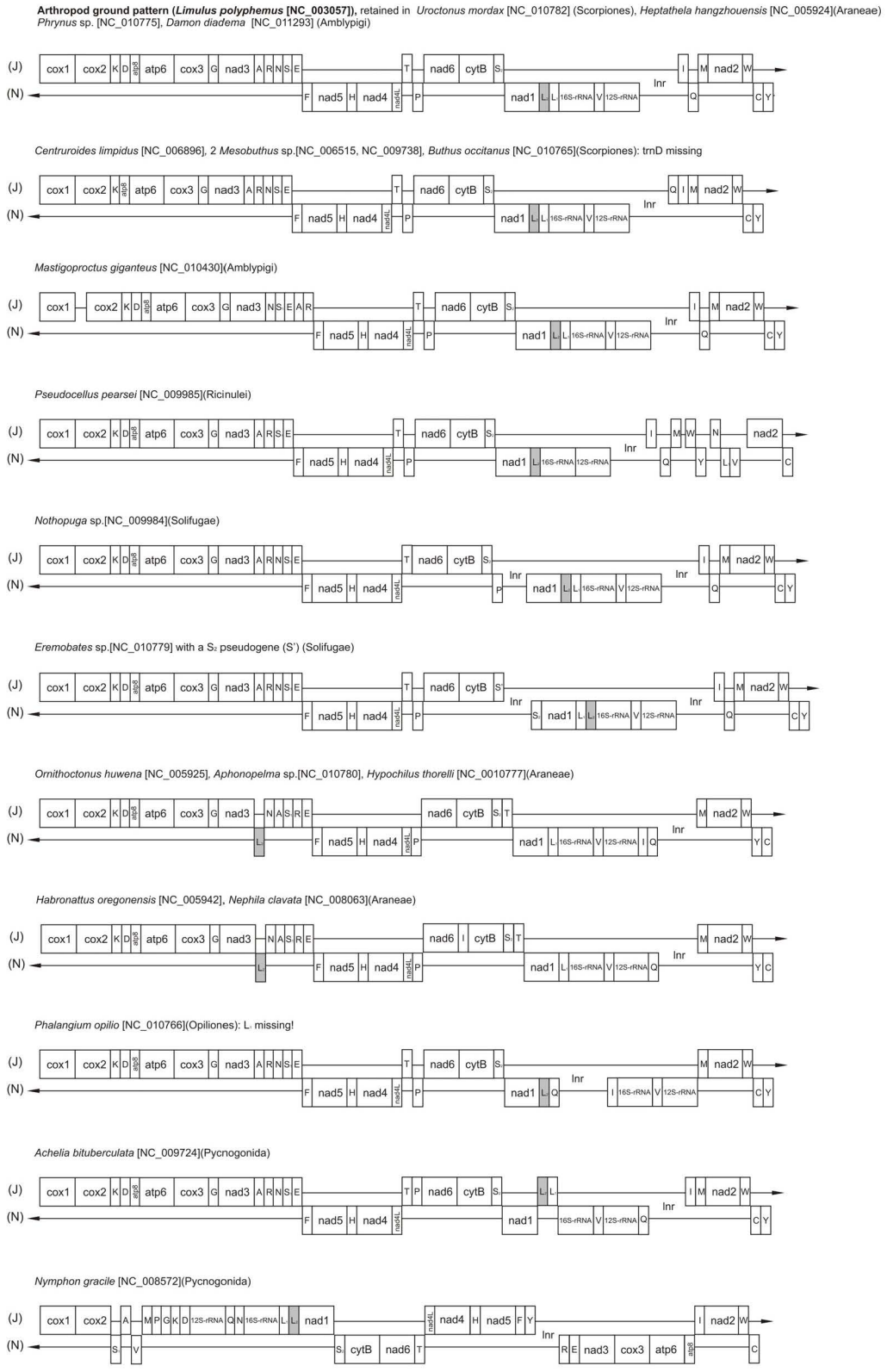
mite and tick (mutant) lines, mite cell lines) will be needed to get Acari (post-)genomics exponentially boosted. In addition, with more (>10) mite and tick genomes to be completely sequenced and the establishment of NGS technologies for expression profiling, a solid bioinformatics expertise will be required to analyze the tsunami of data that will be generated. This new type of Acari research will thus be quite challenging but will, given the enormous diversity of mite and tick species, never be less than fascinating.

Appendices

Chapter II



Appendix II-A



Appendix II-A (continued)- Mitochondrial genome arrangements of 42 Chelicerata

Graphical linearisation of the mt genomes was done according to Fahrein et al. 2003 (see Fig. 3, Chapter II). Corresponding GenBank accession numbers are between brackets. The position of $trnL_2$ is grey shaded. Protein coding and rRNA genes are abbreviated as in the Abbreviations section; tRNA genes are abbreviated using the one-letter amino acid code. Small non-coding regions (> 50bp) are indicated as gaps between genes. Braces accentuate the duplicated region in the mt genome of *M. occidentalis*.

Available online at http://www.biomedcentral.com/1471-2164/10/107/additional

Appendix II-B - Start and stop codons of mt protein coding genes of complete mt genomes of Acari (.xls)

Available online at http://www.biomedcentral.com/1471-2164/10/107/additional

Appendix II-C - GenBank accession numbers and sequences of ESTs of *D. pteronyssinus* and *D. farinae* covering the mt genome of *D. pteronyssinus* (.fasta)

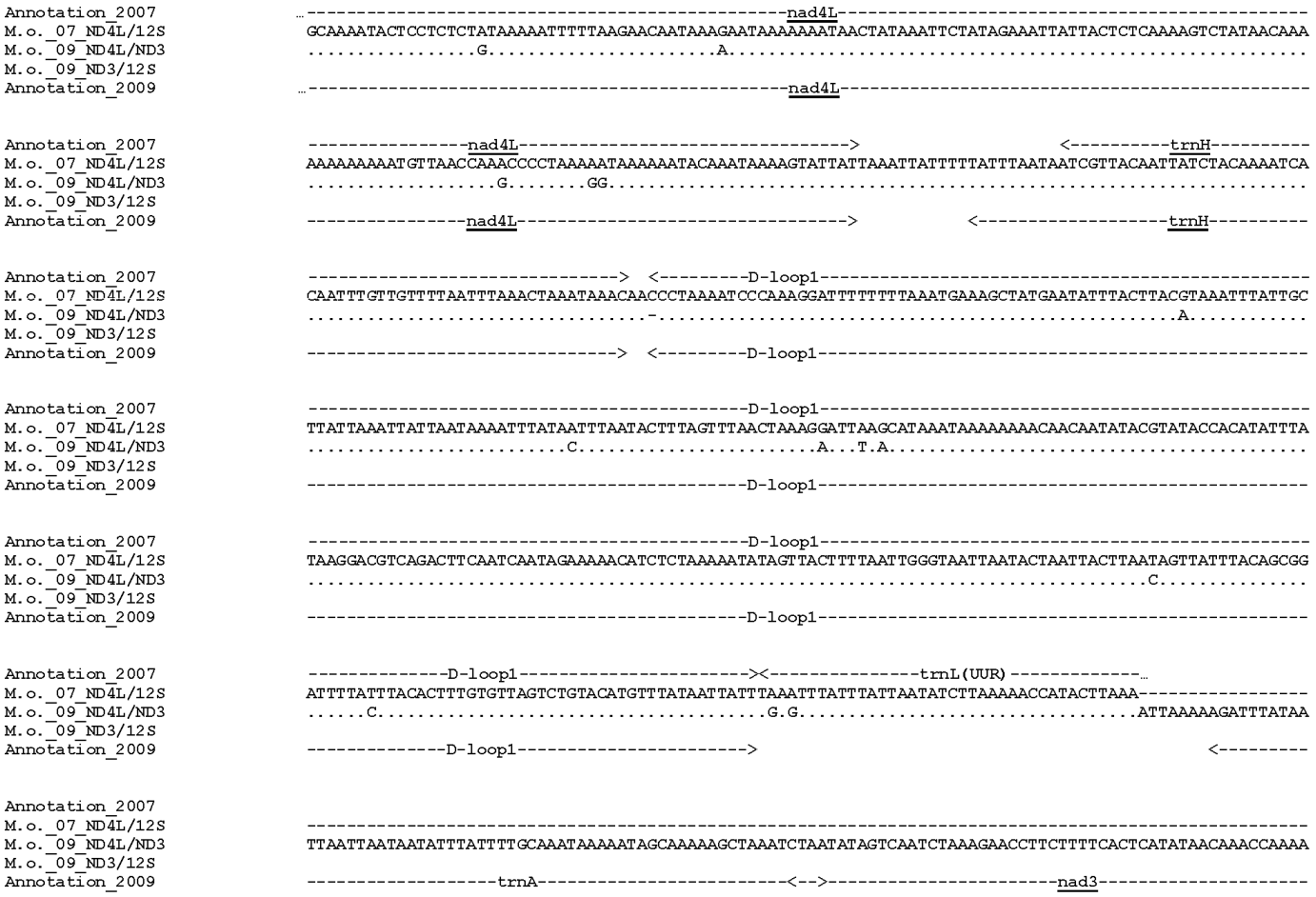
Available online at http://www.biomedcentral.com/1471-2164/10/107/additional

(best viewed with BioEdit 7.0.1 (http://www.mbio.ncsu.edu/BioEdit/BioEdit.html))

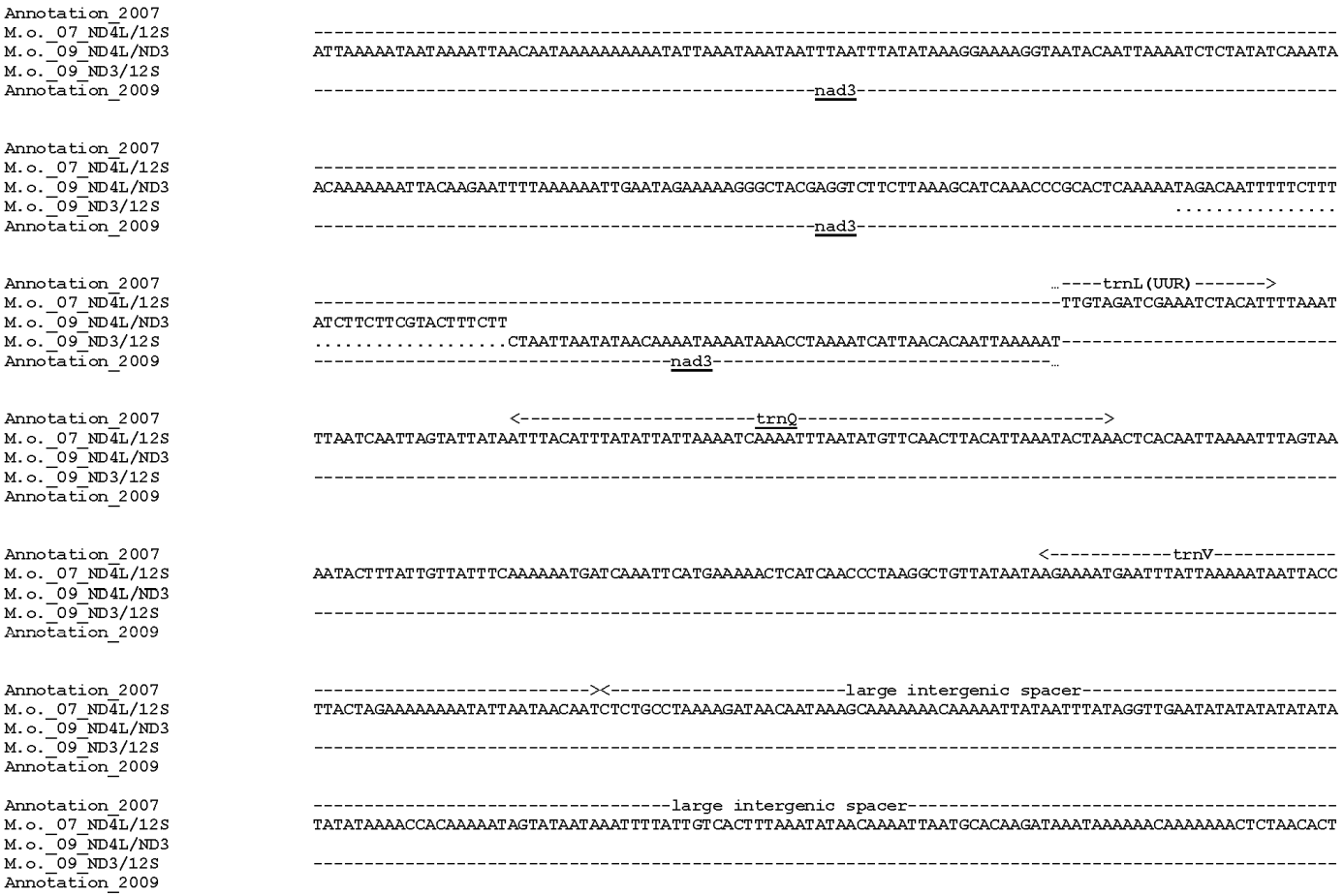
Appendix II-D - Input data for CREx (.xls)

Available online at http://www.biomedcentral.com/1471-2164/10/107/additional

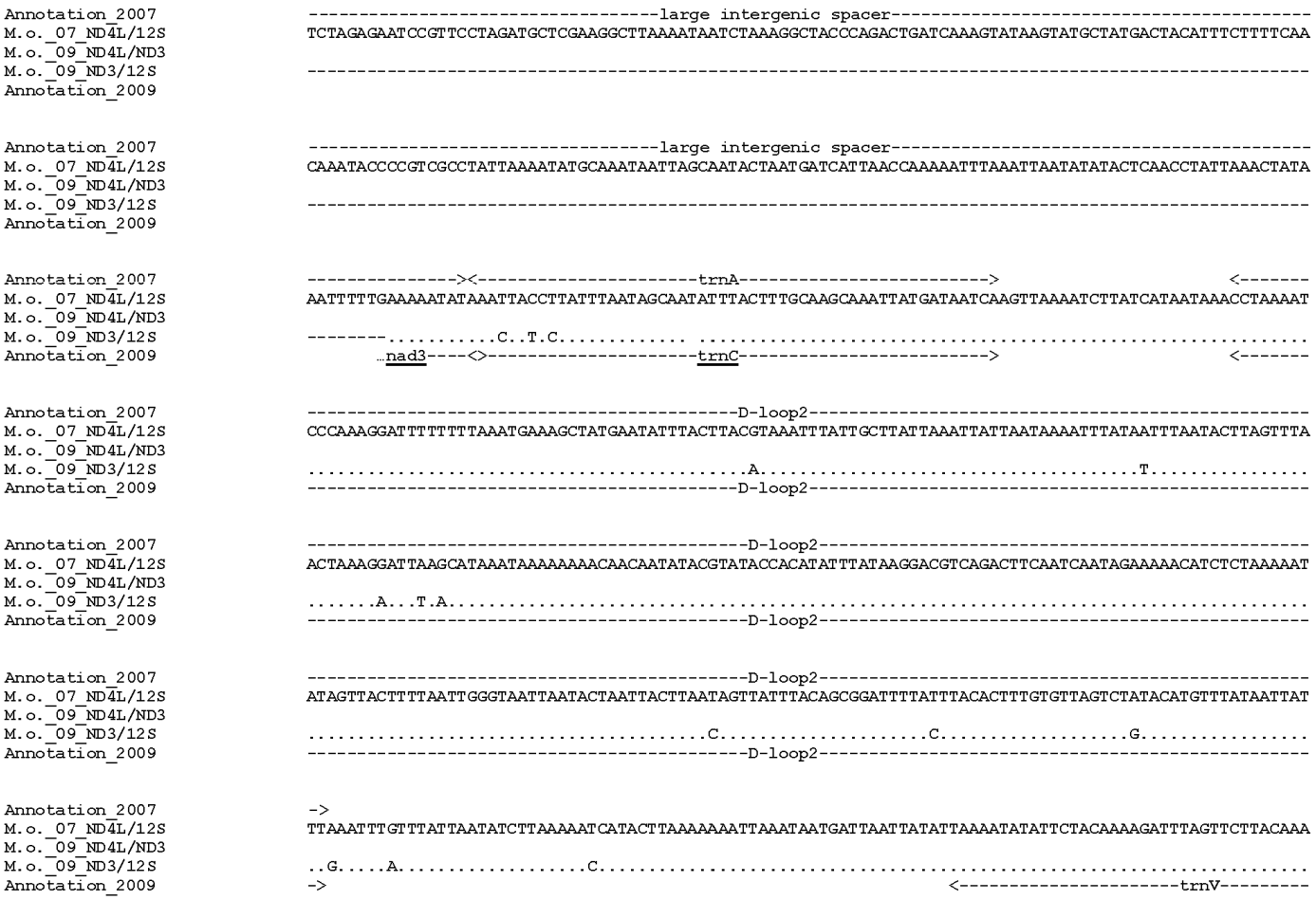
Chapter III



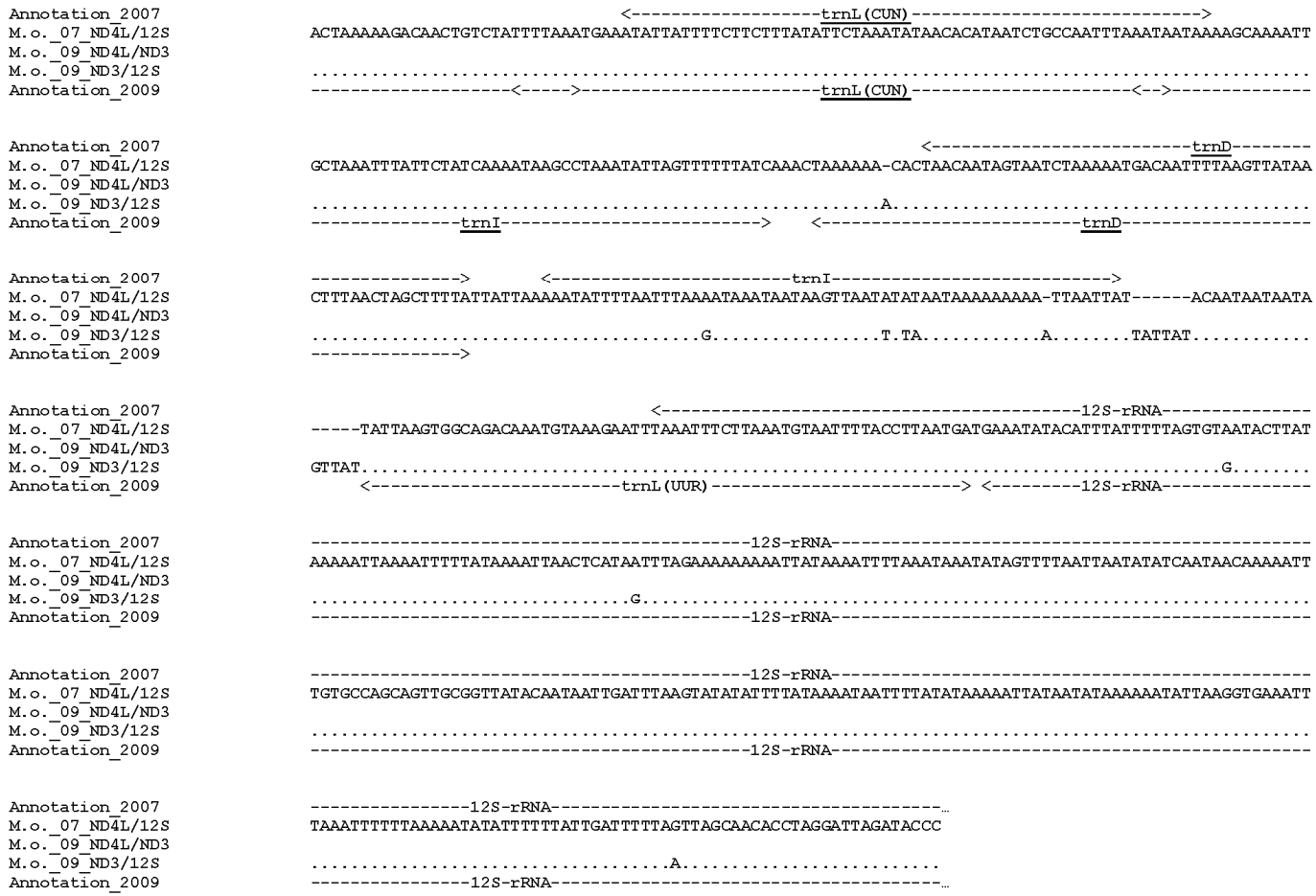
Appendix III-A



Appendix III-A (continued)



Appendix III-A (continued)



Appendix III-A (continued)– Alignment of the *nad4L*/*12S-rRNA* region of the mt genome of *M. occidentalis*.

All sequences are in the 5'-3' direction. Annotation of the *nad4L*/*12S-rRNA* region according to Jeyaprakash and Hoy (2007) (M.o._07_ND4L/12S, GenBank: NC_009093) is positioned above the sequences (Annotation_2007) while annotation of the two overlapping fragments (M.o._09_ND4L/ND3 and M.o._09_ND3/12S, GenBank: GU066253), containing the *nad3* gene (this work), is positioned below the sequences (Annotation_2009). Underlined genes are transcribed on the reverse complementary strand.

Chapter IV

Appendix IV-A - GenBank accession numbers of ID-RCDs used for phylogenetic analysis.

***Tetranychus* sp. ID-RCDs**

Genus[1]	Acc. number[2]	S_p[3]	Genus[1]	Acc. number[2]	S_p[3]	Genus[1]	Acc. number[2]	S_p[3]
Tetranychus urticae	tetur01g00490	Y	*Tetranychus urticae*	tetur07g05940	Y	*Tetranychus urticae*	tetur20g01790	Y
Tetranychus urticae	tetur04g00150	Y	*Tetranychus urticae*	tetur07g06560	Y	*Tetranychus urticae*	tetur28g01250	Y
Tetranychus urticae	tetur04g08620	Y	*Tetranychus urticae*	tetur12g04761	Y	*Tetranychus urticae*	tetur44g00140	Y
Tetranychus urticae	tetur06g00450	Y	*Tetranychus urticae*	tetur13g04550	Y	*Tetranychus evansi*	JQ736355	
Tetranychus urticae	tetur06g00460	Y	*Tetranychus urticae*	tetur19g02300	Y	*Tetranychus evansi*	JQ736356	
Tetranychus urticae	tetur07g02040	Y	*Tetranychus urticae*	tetur19g03360	Y	*Tetranychus evansi*	JQ736357	
Tetranychus urticae	tetur07g05930	Y	*Tetranychus urticae*	tetur20g01160	Y	*Tetranychus evansi*	JQ736358	
						Tetranychus evansi	JQ736359	

"classical" ID-RCDs

Genus[1]	Acc. number[2]	S_p[3]	Genus[1]	Acc. number[2]	S_p[3]	Genus[1]	Acc. number[2]	S_p[3]
Acinetobacter	Q1WCN0	N	*Chelativorans*	Q11GN2	N	*Pyrenophora*	B2WPE2	N
Agrobacterium	Q7CV76	N	*Escherichia*	B7LQX6	N	*Ralstonia*	Q0KAA6	N
Ajellomyces	C5JTU1	N	*Jannaschia*	Q28JN2	N	*Rhodococcus*	Q6F4M7	N
Arthrobacter	A1RBV0	N	*Pandoraea*	A7KX07	N	*Streptomyces*	Q9APK9	N
Arthrobacter	Q76CC4	N	*Paracoccidioides*	C1GHP3	N	*Talaromyces*	B8M1V8	N
Aspergillus	B8NN75	N	*Pseudomonas*	A6V876	N			
Candidatus	Q02CF8	N	*Pyrenophora*	B2VWY2	N			

NCBI BLASTp hits (e-value ≤ e-10) using *T. urticae* ID-RCDs as query

Genus[1]	Acc. number[2]	S_p[3]	Genus[1]	Acc. number[2]	S_p[3]	Genus[1]	Acc. number[2]	S_p[3]
Acaryochloris	YP_001515498	N	*Frankia*	YP_479231	N	*Phytophthora*	XP_002905783	Y
Acidobacterium	ZP_07065922	Y	*Frankia*	YP_001508082	N	*Podospora*	CAP65306	Y
Acidobacterium	ZP_07648583	N	*Frankia*	YP_001508663	N	*Podospora*	XP_001903249	Y
Actinosynnema	YP_003100310	N	*Frankia*	YP_004019605	N	*Pseudomonas*	YP_350088	N
Actinosynnema	YP_003100966	N	*Frankia*	ZP_06410955	N	*Puccinia*	EFP84436	Y
Ajellomyces	EEH02815	Y	*Frankia*	ZP_06475430	N	*Pyrenophora*	EFQ91909	Y
Ajellomyces	EER41332	Y	*Fulvimarina*	ZP_01438862	Y	*Pyrenophora*	XP_001940623	Y
Ajellomyces	EEQ89061	Y	*Gemmatimonas*	YP_002762722	N	*Ralstonia*	YP_728952	N
Alteromonas	ZP_04713930	N	*Gibberella*	XP_382158	Y	*Ralstonia*	YP_299431	N
Amycolatopsis	YP_003765631	Y	*Gibberella*	XP_384861	Y	*Rhizobium*	ZP_03505301	Y
Arthrobacter	YP_949837	Y	*Gibberella*	XP_388213	Y	*Rhizobium*	YP_470103	Y
Arthrobacter	YP_002489952	N	*Glomerella*	EFQ27291	Y	*Rhizobium*	YP_001978859	Y
Arthrobacter	YP_833604	Y	*Glomerella*	EFQ34262	N	*Rhizobium*	ZP_03530195	Y
Aspergillus	CBF87292	N	*Glomerella*	EFQ35214	Y	*Rhizobium*	ZP_03524635	N
Aspergillus	XP_658905	Y	*Haloferax*	YP_003536123	N	*Rhodococcus*	YP_004008885	Y
Aspergillus	XP_661807	Y	*Herpetosiphon*	YP_001546287	Y	*Rhodococcus*	YP_004008886	Y
Aspergillus	XP_662932	Y	*Herpetosiphon*	YP_001547597	N	*Rhodococcus*	ZP_06830704	Y
Aspergillus	XP_663002	Y	*Hyphomicrobium*	YP_003756407	N	*Rhodococcus*	ZP_06830705	Y
Aspergillus	XP_664765	Y	*Janibacter*	ZP_00995526	N	*Rhodococcus*	AAN37493	N
Aspergillus	XP_001389482	N	*Janibacter*	ZP_00995926	N	*Rhodococcus*	AAR27826	Y
Aspergillus	XP_001389508	Y	*Ketogulonicigenium*	YP_003964856	Y	*Riemerella*	YP_004044931	N
Aspergillus	XP_001390849	Y	*Kineococcus*	YP_001363667	N	*Roseibium*	ZP_07660688	Y
Aspergillus	XP_001391162	N	*Kitasatospora*	BAJ27524	N	*Roseiflexus*	YP_001431698	Y
Aspergillus	XP_001393826	Y	*Kitasatospora*	BAJ31863	N	*Roseiflexus*	YP_001275304	Y
Aspergillus	XP_001395585	N	*Kribbella*	YP_003381075	N	*Saccharopolyspora*	YP_001104202	Y
Aspergillus	XP_001395793	Y	*Laccaria*	XP_001879555	Y	*Saccharopolyspora*	YP_001105681	Y
Aspergillus	XP_001396120	Y	*Limnobacter*	ZP_01915839	Y	*Sanguibacter*	YP_003314415	N
Aspergillus	XP_001399989	Y	*Magnaporthe*	XP_363006	Y	*Schistosoma*	XP_002569539	N
Aspergillus	CAK37271	Y	*Marinomonas*	ZP_01076090	Y	*Schizophyllum*	XP_003026007	Y
Aspergillus	CAK40367	Y	*Maritimibacter*	ZP_01013228	Y	*Schizophyllum*	XP_003026084	Y
Aspergillus	CAK41421	Y	*Meiothermus*	YP_003505838	N	*Schizophyllum*	XP_003029142	N
Aspergillus	CAK44937	Y	*Mesorhizobium*	YP_004139576	Y	*Schizophyllum*	XP_003029311	N
Aspergillus	CAK46228	Y	*Mesorhizobium*	YP_675384	Y	*Schizophyllum*	XP_003031024	Y
Asticcacaulis	YP_004087964	N	*Methylobacterium*	YP_002423474	Y	*Schizophyllum*	XP_003035277	N
Beutenbergia	YP_002880947	N	*Methylosinus*	ZP_06887255	Y	*Sclerotinia*	XP_001584708	Y
Botryotinia	XP_001547782	N	*Methylovorus*	YP_004039767	Y	*Sclerotinia*	XP_001590335	Y
Botryotinia	XP_001550224	Y	*Methylovorus*	YP_003051066	Y	*Sinorhizobium*	YP_002823995	Y
Bradyrhizobium	YP_001243151	Y	*Micrococcus*	YP_002956532	N	*Sinorhizobium*	YP_001312810	Y
Bradyrhizobium	YP_001202885	Y	*Micrococcus*	ZP_06500982	N	*Sorangium*	YP_001617686	N
Brevibacterium	ZP_05914095	N	*Micromonospora*	ZP_04603856	N	*Sorangium*	YP_001617690	N
Burkholderia	ZP_03270367	N	*Moniliophthora*	XP_002390944	N	*Spirosoma*	YP_003387063	Y
Caulobacter	YP_003594737	Y	*Mucilaginibacter*	ZP_07746550	Y	*Stigmatella*	ZP_01459590	Y
Cellvibrio	YP_001983304	N	*Naegleria*	XP_002681734	Y	*Streptomyces*	YP_003490106	N
Chaetomium	XP_001223092	Y	*Nakamurella*	YP_003203127	N	*Streptomyces*	YP_003490270	Y
Chaetomium	XP_001224578	Y	*Nectria*	XP_003040987	Y	*Streptomyces*	YP_003490910	N
Chaetomium	XP_001228656	Y	*Nectria*	XP_003043843	Y	*Streptomyces*	ZP_06275428	N
Chitinophaga	YP_003120875	Y	*Nectria*	XP_003047595	Y	*Streptomyces*	ZP_07284524	N
Chloroflexus	YP_002463143	N	*Nectria*	XP_003052962	Y	*Streptomyces*	ZP_07285769	N
Chloroflexus	YP_001635170	N	*Neosartorya*	XP_001257499	Y	*Streptomyces*	ZP_06708934	N
Clavibacter	YP_001221487	N	*Neosartorya*	XP_001258835	Y	*Streptomyces*	ZP_06712030	N
Comamonas	YP_003280206	N	*Neosartorya*	XP_001263001	Y	*Streptomyces*	ZP_04999238	Y
Comamonas	ZP_03541192	N	*Neosartorya*	XP_001265559	Y	*Streptomyces*	ZP_06916827	N

Appendix IV-A (continued)

NCBI BLASTp hits (e-value ≤ e-10) using *T. urticae* ID-RCDs as query

Genus[1]	Acc. number[2]	S_p[3]	Genus[1]	Acc. number[2]	S_p[3]	Genus[1]	Acc. number[2]	S_p[3]
Comamonas	ZP_07046125	N	*Neurospora*	XP_963883	Y	*Streptomyces*	ZP_06918248	Y
Conexibacter	YP_003395019	N	*Nitrosococcus*	YP_003527272	N	*Streptomyces*	ZP_06918456	N
Coprinopsis	XP_001831803	Y	*Nitrosococcus*	YP_343162_Y	Y	*Streptomyces*	ZP_06918464	Y
Coprinopsis	XP_001834886	Y	*Nitrosococcus*	YP_344402	N	*Streptosporangium*	YP_003343260	N
Corynebacterium	NP_599490	N	*Nitrosospira*	YP_412780	Y	*Talaromyces*	XP_002478049	Y
Corynebacterium	YP_001137177	N	*Nocardia*	YP_120510	N	*Talaromyces*	XP_002488138	Y
Corynebacterium	ZP_07404061	Y	*Novosphingobium*	YP_497150	N	*Thermobaculum*	YP_003323670	N
Corynebacterium	ZP_07990861	N	*Paracoccidioides*	XP_002796777	Y	*Thermus*	ZP_03495684	Y
Cryptococcus	XP_572061	Y	*Paracoccidioides*	EEH19263	Y	*Thermus*	YP_005116	N
Cupriavidus	YP_587403	Y	*Paracoccidioides*	EEH48111	Y	*Thiobacillus*	YP_314960	N
Cupriavidus	YP_002007309	Y	*Pectobacterium*	YP_003019323	Y	*Tuber*	XP_002840547	Y
Debaryomyces	XP_002770542	Y	*Pelagibaca*	ZP_01444072	N	*Ustilago*	XP_760258	Y
Deinococcus	YP_002784902	N	*Penicillium*	XP_002556982	Y	*Ustilago*	XP_762135	Y
Deinococcus	DYP_002787243	N	*Penicillium*	XP_002559399	Y	*Verticillium*	XP_003005923	Y
Deinococcus	YP_594344	N	*Penicillium*	XP_002563845	Y	*Xenorhabdus*	YP_003467843	Y
Delftia	YP_001564166	N	*Phaeosphaeria*	XP_001791242	Y	*Xenorhabdus*	YP_003712511	Y
Dyadobacter	YP_003088138	Y	*Phaeosphaeria*	XP_001791707	Y	*Yersinia*	NP_670676	N
Flavobacteriales	ZP_02181399	Y	*Photorhabdus*	YP_003042719	N	*Yersinia*	YP_001399705	N
Flavobacterium	YP_001194031	Y	*Photorhabdus*	NP_931592	Y	*Yersinia*	YP_071754	N

[1] *T. urticae* accession numbers can be accessed through the ORCAE genome portal (http://bioinformatics.psb.ugent.be/orcae/overview/Tetur), other accession (acc.) numbers are available at GenBank

[2] colour code: bacteria: green; fungi: purple; Archaea: red; Chromalveolata: dark blue; Excavata: light blue; Animalia: yellow

[3] S_p: predicted with (Y) or without a (N) signal peptide according to SignalP 3.0 (Bendtsen *et al*. 2004)

Appendix IV-B - Accession numbers of lipocalins used for phylogenetic analysis

Species	Taxonomy	Name	Acc. number
Tetranychus urticae (58)	Arachnida: Acari: Trombidiformes	-	see Table IV.3
Ahrensia sp.	Alphaproteobacteria: Rhodobacteriales	outer membrane lipoprotein Blc	ZP_07375698.1
Homarus gammarus	Crustacea: Decapoda	Crustacyanin-A1	P58989.1
Homarus gammarus	Crustacea: Decapoda	Crustacyanin-A2	P80007.1
Debaromyces hansenii	Fungi: Ascomycota	-	XP_460369.1
Klebsiella pneumoniae	Proteobacteria: Enterobacteriales	lipoprotein Blc	ZP_06018160.1
Anopheles gambiae	Insecta: Diptera	Agam.Lip1	XP_320076.4
Drosophila melanogaster	Insecta: Diptera	neural lazarillo	AAF51378.2
Drosophila melanogaster	Insecta: Diptera	glial lazarillo	NP_523727.2
Drosophila melanogaster	Insecta: Diptera	Karl	AAK72697.1
Rhodnius prolixus	Insecta: Hemiptera	biogenic amine binding protein	AAO25746.1
Rhodnius prolixus	Insecta: Hemiptera	*Rhodnius* platelet aggregation inhibitor 3	AAQ20817.1
Rhodnius prolixus	Insecta: Hemiptera	*Rhodnius* platelet aggregation inhibitor 4	AAQ20818.1
Rhodnius prolixus	Insecta: Hemiptera	Nitrophorin-1	Q26239.1
Rhodnius prolixus	Insecta: Hemiptera	pallidipin salivary platelet aggregation inhibitor 1	AAB09090.1
Rhodnius prolixus	Insecta: Hemiptera	triabin-like lipocalin 1	AAQ20821.1
Triatoma brasiliensis	Insecta: Hemiptera	salivary triabin 1	ABH09425.1
Triatoma brasiliensis	Insecta: Hemiptera	pallidipin precursor	ABH09434.1
Triatoma pallidipennis	Insecta: Hemiptera	pallidipin 2	AAA30329.1
Triatoma pallidipennis	Insecta: Hemiptera	triabin	CAA56540.1
Apis mellifera	Insecta: Hymenoptera	Amel_Lip1	XP_392555.1
Apis mellifera	Insecta: Hymenoptera	Amel_Lip2	XP_623787.2
Harpegnathos saltator	Insecta: Hymenoptera	apolipoprotein D	EFN80483.1
Bombyx mori	Insecta: Lepidoptera	Bombyrin	NP_001036872.1
Bombyx mori	Insecta: Lepidoptera	32 kDa apolipoprotein	NP_001140192.1
Galleria mellonella	Insecta: Lepidoptera	gallerin	AAA85089.1
Hyphantria cunea	Insecta: Lepidoptera	hyphantrin	AAM18117.2
Lonomia obliqua	Insecta: Lepidoptera	lipocalin 1 (biliprotein)	AAV91447.1
Lonomia obliqua	Insecta: Lepidoptera	lipocalin 4 (biliprotein)	AAV91423.1
Manduca sexta	Insecta: Lepidoptera	Insecticyanin 1	P00305.1
Manduca sexta	Insecta: Lepidoptera	Insecticyanin 2	Q00630.1
Pieris brassicae	Insecta: Lepidoptera	bilin-binding protein	CAA54063.1
Samia cynthia ricini	Insecta: Lepidoptera	biliverdin binding protein-I	BAB85482.1
Samia cynthia ricini	Insecta: Lepidoptera	biliverdin binding protein-II	BAB84676.1
Shistocerca americana	Insecta: Orthoptera	Lazarillo	CAA86216.1
Arabidopsis thaliana	Plantae: Brassicaceae	-	AAM62904.1
Zea mays	Plantae: Poaceae	-	NP_001140887.1
Oryza sativa	Plantae: Poaceae	-	NP_001047416.1
Bos taurus	Vertebrata: Mammalia	apolipoprotein D	DAA33372.1
Homo sapiens	Vertebrata: Mammalia	apolipoprotein D	NP_001638.1
Mus musculus	Vertebrata: Mammalia	apolipoprotein D	CAA57974.1

Chapter V

Appendix V-A - Alignment of N-terminal NBDs of *T. urticae* ABC proteins (ABC_N-term_NBD_Tetranychus.fas), N- and C-terminal NBDs of metazoan ABC proteins (ABC_NBD_Metazoa_simple.fas) and of full length metazoan ABCA, -B, C, D, E, F, G and H protein sequences (ABCA-H.fas) (bundled in a .rar file)
Available online at: http://www.biomedcentral.com/bmcgenomics
(best viewed with BioEdit 7.0.1 (http://www.mbio.ncsu.edu/BioEdit/BioEdit.html))

Appendix V-B - Amino acid substitution models used for maximum likelihood phylogenetic analyses and likelihood scores of constructed phylogenetic trees (.xlsx).
Available online at: http://www.biomedcentral.com/bmcgenomics

Appendix V-C - Accession numbers of sequences used for phylogenetic analysis (.xlsx).
Available online at: http://www.biomedcentral.com/bmcgenomics

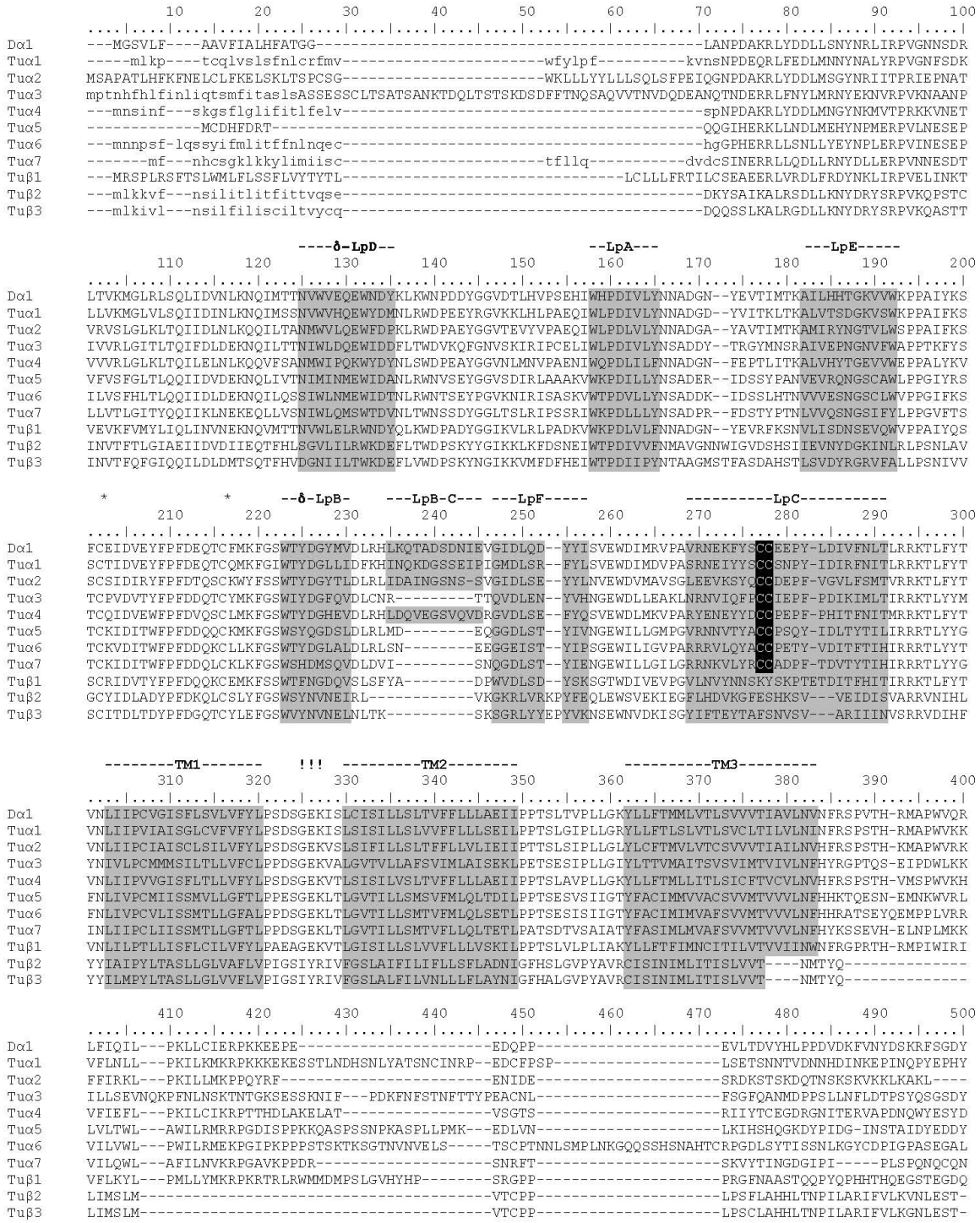
Appendix V-D - Amino acid sequences of 103 *T. urticae* ABCs (.fasta)
Available online at: http://www.biomedcentral.com/bmcgenomics
(best viewed with BioEdit 7.0.1 (http://www.mbio.ncsu.edu/BioEdit/BioEdit.html))

Appendix V-E - Exon-intron pattern of 103 *T. urticae* ABCs (.pdf)
Exons are depicted as light grey boxes while strandlines represent introns. Within exons red boxes represent NBDs, while small dark grey boxes represent transmembrane helices (TM). Solid arrows indicate direction of transcription Available online at: http://www.biomedcentral.com/bmcgenomics

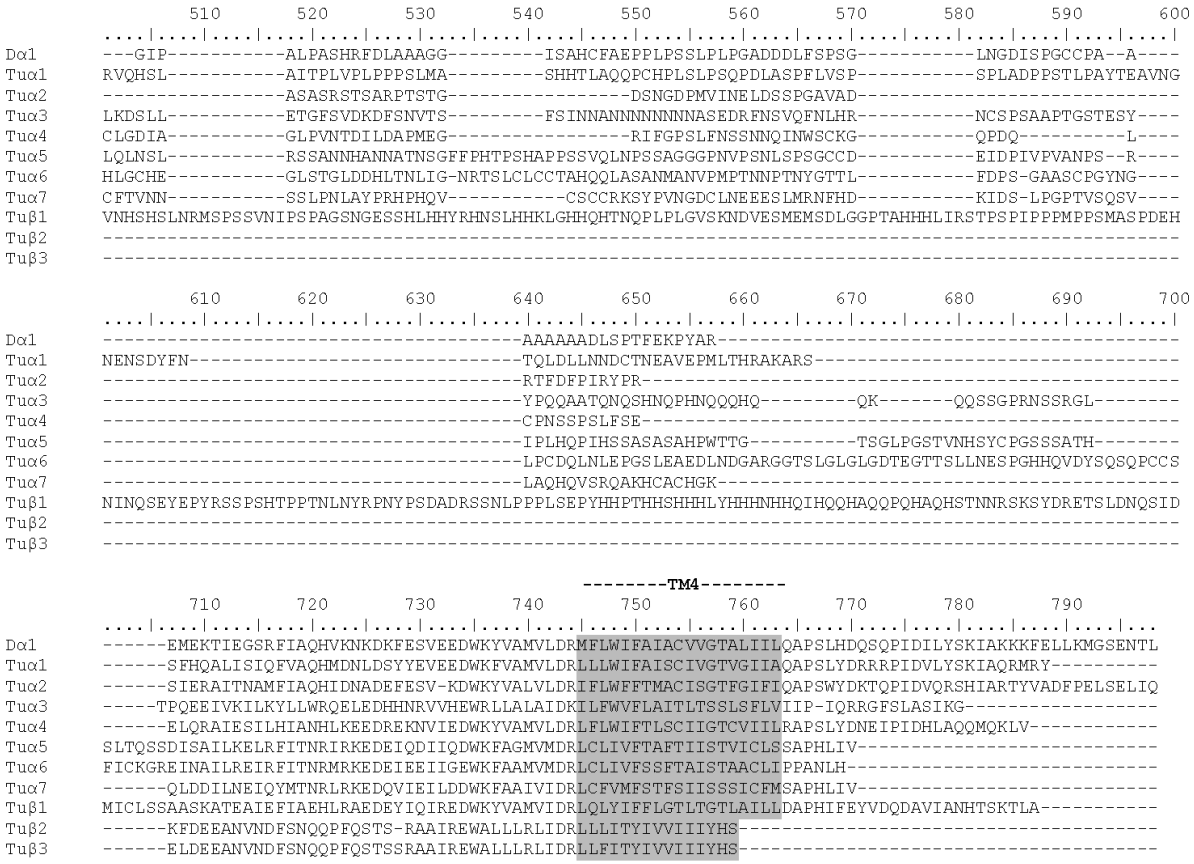
Appendix V-F - Identity/similarity matrices between *T. urticae, D. melanogaster* and *H. sapiens* ABC proteins (.xlsx)
In the upper triangle identity values are shown, while in the lower triangle similarity values are presented.
Available online at: http://www.biomedcentral.com/bmcgenomics

Appendix V-G - Fold change values of differentially expressed ABC genes in mites after host plant change to either *Arabidopsis* or tomato (.xlsx).
Available online at: http://www.biomedcentral.com/bmcgenomics

Chapter VI

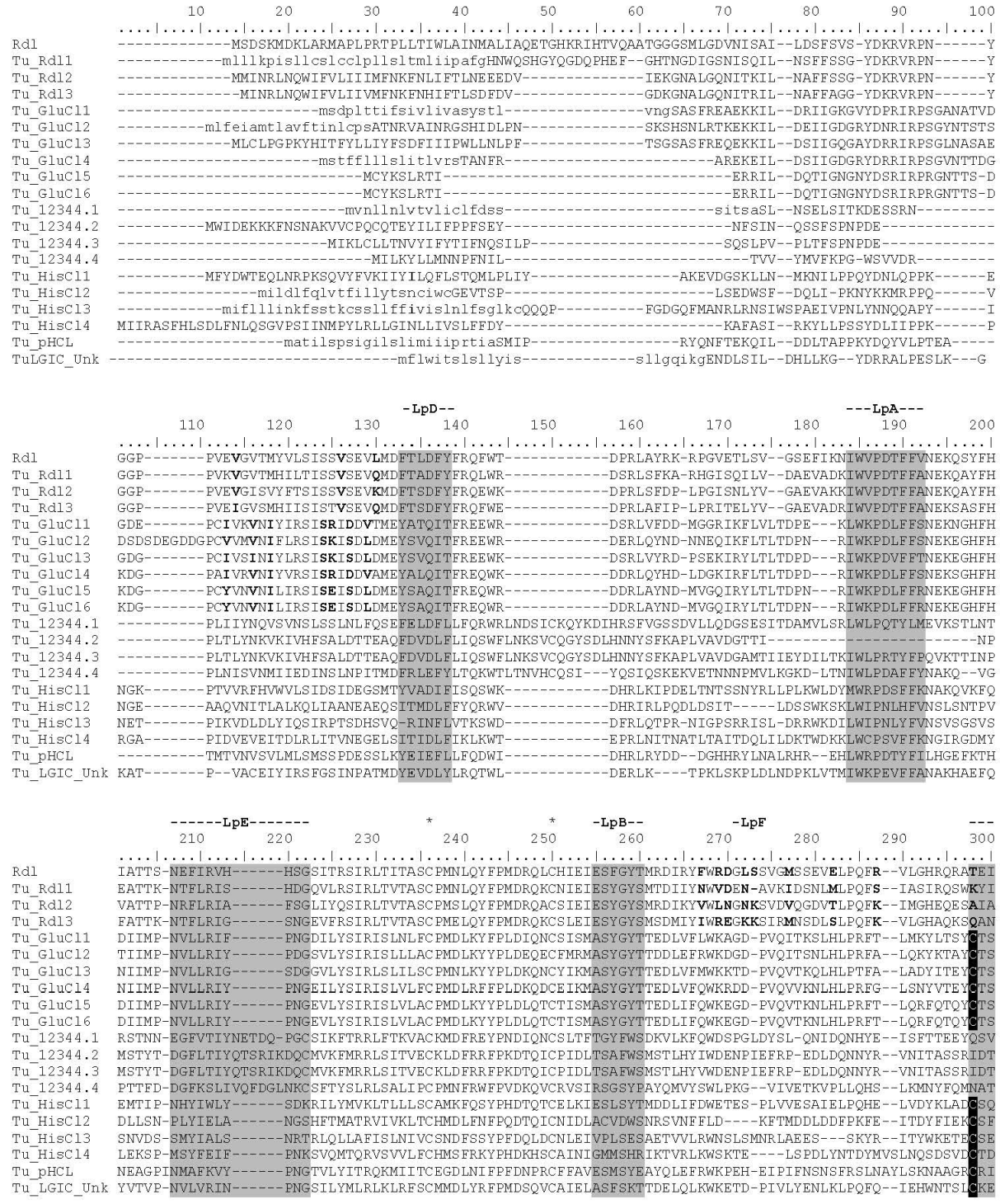


Appendix VI-A

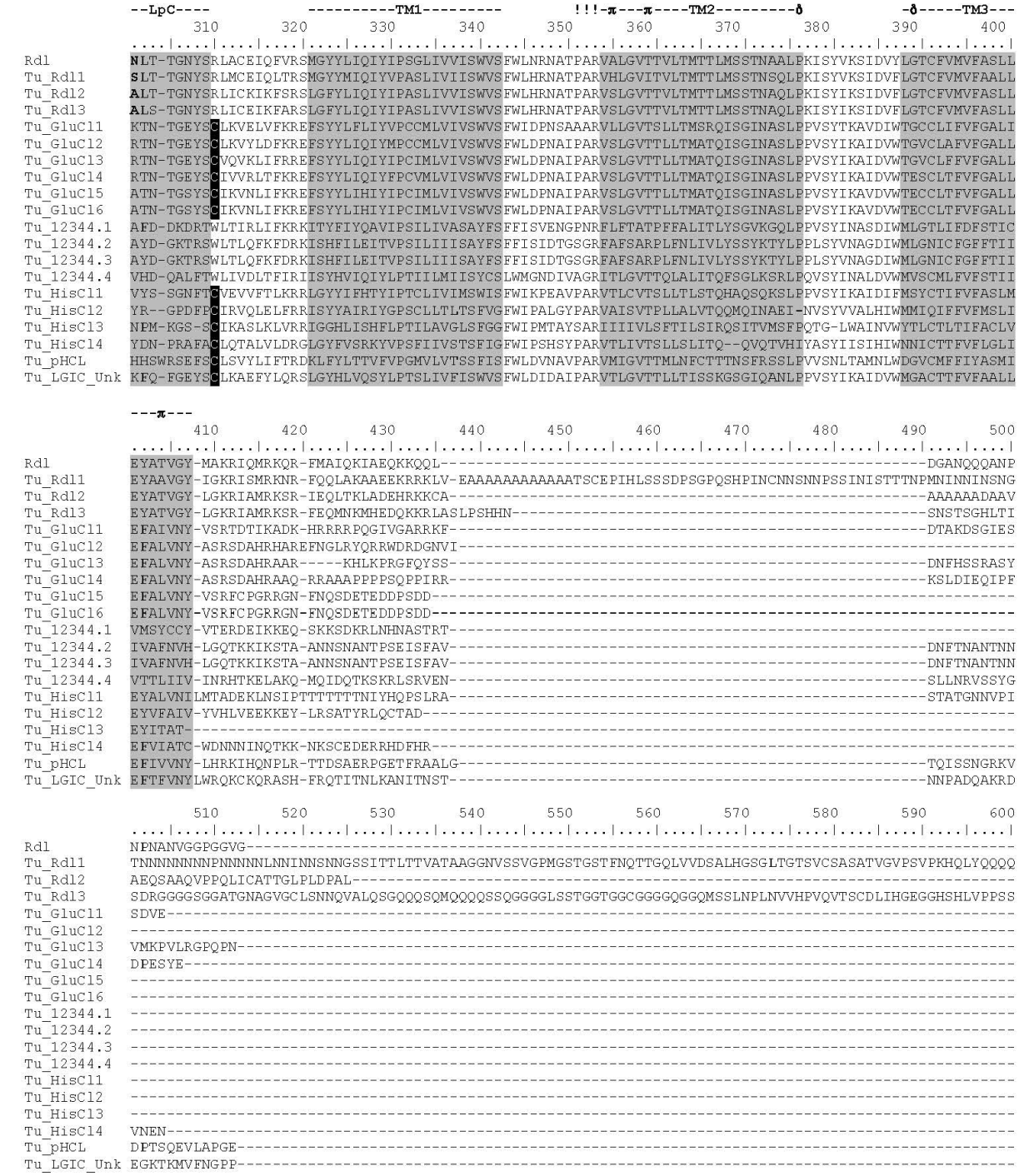


Appendix VI-A (continued) - Protein sequence alignment of *T. urticae* nAChR subunits with Dα1 of *D. melanogaster.*

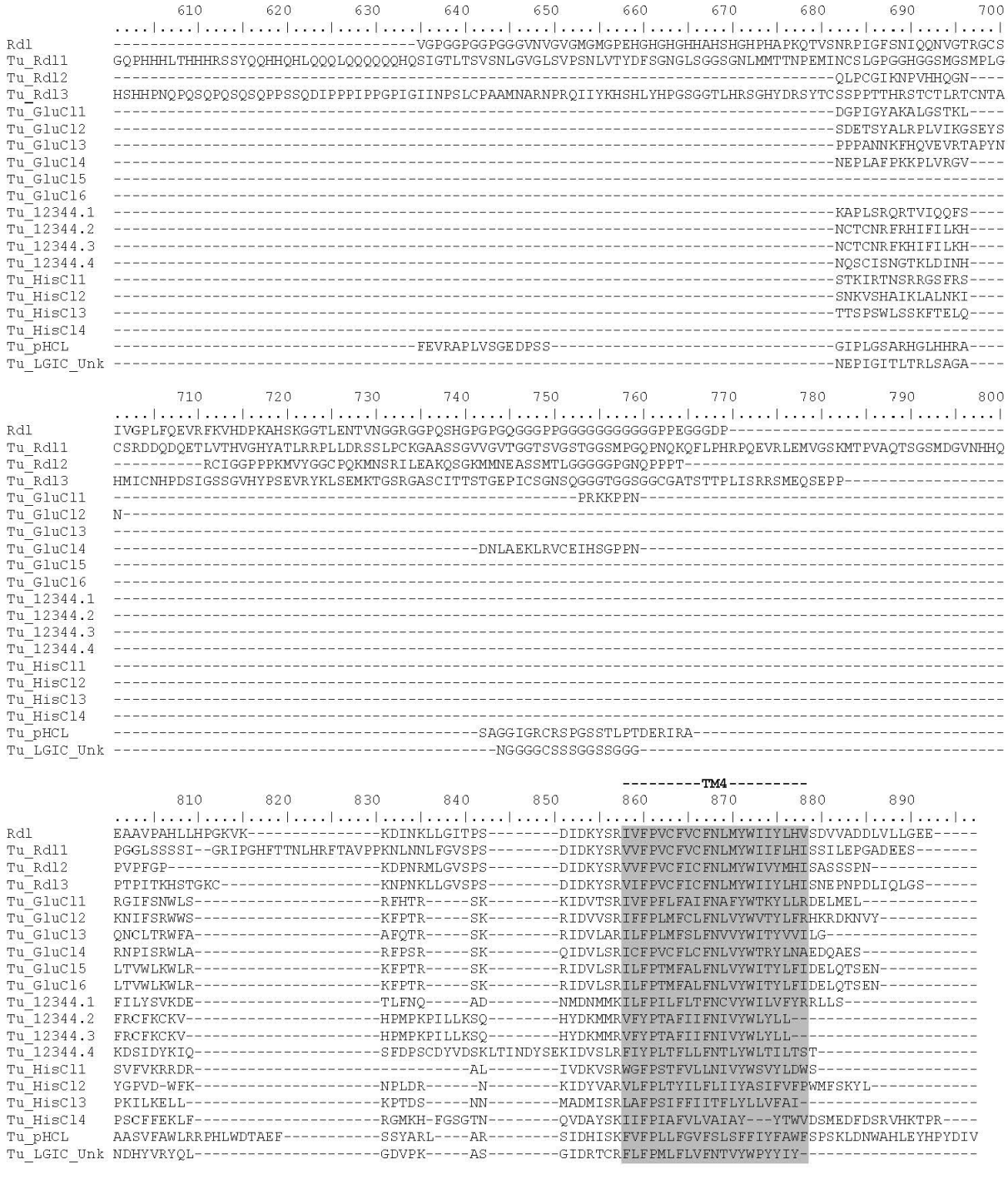
Predicted N-terminal signal peptide leader sequences are shown in lower case and loops implicated in ligand binding (LpA-F) and the four transmembrane regions (TM1-4) are shaded in grey. Residues preceding TM2, important for ion charge selectivity, are indicated by exclamation marks. The two cysteines forming the cys-loop are marked by asterisks. The vicinal cysteines, typical for α subunits are shaded in black. Positions of amino acid substitutions involved in imidacloprid resistance (Y151S (*N. lugens* numbering), R81T (*M. persicae* numbering)) in insects are marked with a δ symbol. Available online at: http://www.sciencedirect.com/science/article/pii/S0965174812000410



Appendix VI-B



Appendix VI-B (continued)



Appendix VI-B (continued)

Appendix VI-B (continued) - Protein sequence alignment of *T. urticae* cysLGIC subunits other than nAChRs with Rdl of *D. melanogaster*

Predicted N-terminal signal peptides are shown in lower case and loops implicated in ligand binding (LpA-F) and the four transmembrane regions (TM1-4) are shaded in grey. Residues preceding TM2, important for ion charge selectivity, are indicated by exclamation marks. The two cysteines forming the cys-loop are marked by asterisks, while 2 additional cysteines (Dent, 2006) are highlighted by black shading. Substitutions associated with ivermectin (G314D/G326E (*T. urticae* numbering) and P299S (*D. melanogaster* numbering)) and dieldrin/fipronil resistance (A301S, T305L and T350M (*D. melanogaster* numbering)) in arthropods are marked with a δ and π symbol, respectively (Buckingham *et al.*, 2005; Hope *et al.*, 2010; Kane *et al.*, 2000; Kwon *et al.*, 2010; Nakao *et al.*, 2010). Positions of amino acids that are not conserved in alternative spliced exons (e.g. between *D. melanogaster* RDL exon 3a and exon 3b) of cysLGICs of insects (Jones *et al.*, 2010; Jones and Sattelle, 2007) are indicated in bold. Available online at: http://www.sciencedirect.com/science/article/pii/S0965174812000410

Appendix VI-C - Percentage identity/similarity between *T. urticae* and *D. melanogaster* cysLGIC other than nAChR protein sequences (.xlsx)

Available online at: http://www.sciencedirect.com/science/article/pii/S0965174812000410

References

Abascal, F., Zardoya, R., *et al.* (2005). ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21: 2104-2105.

Abascal, F., Posada, D., *et al.* (2006). Parallel evolution of the genetic code in arthropod mitochondrial genomes. PLoS Biology 4: 711-718.

Abascal, F., Posada, D., *et al.* (2007). MtArt: A new model of amino acid replacement for arthropoda. Molecular Biology and Evolution 24: 1-5.

Abascal, F., Posada, D., *et al.* (2012). The evolution of the mitochondrial genetic code in arthropods revisited. Mitochondrial DNA 23: 84-91.

Abbott, W.S. (1925). A method of computing the effectiveness of an insecticide. Journal of Economic Entomology 18: 265-267.

Abo-Elghar, G.E., Fujiyoshi, P., *et al.* (2004). Significance of the sulfonylurea receptor (SUR) as the target of diflubenzuron in chitin synthesis inhibition in *Drosophila melanogaster* and *Blattella germanica*. Insect Biochemistry and Molecular Biology 34: 743-752.

Adams, M.D., Celniker, S.E., *et al.* (2000). The genome sequence of *Drosophila melanogaster*. Science 287: 2185-2195.

Aebersold, R. and Mann, M. (2003). Mass spectrometry-based proteomics. Nature 422: 198-207.

Agrawal, A.A. (2000). Host-range evolution: Adaptation and trade-offs in fitness of mites on alternative hosts. Ecology 81: 500-508.

Agrawal, A.A., Vala, F., *et al.* (2002). Induction of preference and performance after acclimation to novel hosts in a phytophagous spider mite: adaptive plasticity ? American Naturalist 159: 553-565.

Aguinaldo, A.M.A., Turbeville, J.M., *et al.* (1997). Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387: 489-493.

Ahmad, S. (1986). Enzymatic adaptations of herbivorous insects and mites to phytochemicals. Journal of Chemical Ecology 12: 533-560.

Akiyama, M., Sugiyama-Nakagiri, Y., *et al.* (2005). Mutations in lipid transporter ABCA12 in harlequin ichthyosis and functional recovery by corrective gene transfer. Journal of Clinical Investigation 115: 1777-1784.

Ako, M., Borgemeister, C., *et al.* (2004). Effects of neonicotinoid insecticides on the bionomics of twospotted spider mite (Acari : Tetranychidae). Journal of Economic Entomology 97: 1587-1594.

Akrouh, A., Halcomb, S.E., *et al.* (2009). Molecular biology of K-ATP channels and implications for health and disease. Iubmb Life 61: 971-978.

Albrecht, C. and Viturro, E. (2007). The ABCA subfamily - gene and protein structures, functions and associated hereditary diseases. Pfluegers Archiv European Journal of Physiology 453: 581-589.

Ali, J.G. and Agrawal, A.A. (2012). Specialist versus generalist insect herbivores and plant defense. Trends in Plant Science: 1-10.

Aljamali, M.N., Ramakrishnan, V.G., *et al.* (2009). Microarray analysis of gene expression changes in feeding female and male lone star ticks, *Amblyomma americanum* (L). Archives of Insect Biochemistry and Physiology 71: 236-253.

Allikmets, R., Gerrard, B., *et al.* (1996). Characterization of the human ABC superfamily: Isolation and mapping of 21 new genes using the expressed sequence tags database. Human Molecular Genetics 5: 1649-1655.

Altschul, S.F., Gish, W., *et al.* (1990). Basic Local Alignment Search Tool. Journal of Molecular Biology 215: 403-410.

Altschul, S.F., Madden, T.L., *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25: 3389-3402.

Anaka, M., MacDonald, C.D., *et al.* (2008). The white gene of *Drosophila melanogaster* encodes a protein with a role in courtship behavior. Journal of Neurogenetics 22: 243-276.

Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biology 11: R106.

Anderson, S., Bankier, A.T., *et al.* (1981). Sequence and organization of the human mitochondrial genome. Nature 290: 457-465.

Anderson, C.L., Strope, C.L., *et al.* (2011). SuiteMSA: visual tools for multiple sequence alignment comparison and molecular sequence simulation. BMC Bioinformatics 12: 184.

Angus, A., Ong, S., *et al.* (2004). Sequence tag catalogs of dust mite-expressed genomes. American Journal of Pharmacogenomics 4: 357-369.

Anjard, C., the *Dictyostelium discoideum* Consortium, *et al.* (2002). Evolutionary Analyses of ABC Transporters of *Dictyostelium discoideum*. Eukaryotic Cell 1: 643-652.

Annilo, T., Chen, Z.Q., *et al.* (2006). Evolution of the vertebrate ABC gene family: Analysis of gene birth and death. Genomics 88: 1-11.

Arabi, J., Judson, M.L.I., *et al.* (2012). Nucleotide composition of CO1 sequences in Chelicerata (Arthropoda): detecting new mitogenomic rearrangements. Journal of Molecular Evolution 74: 81-95.

Arensburger, P., Megy, K., *et al.* (2010). Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. Science 330: 86-88.

Arita, M., Suematsu, T., *et al.* (2006). An evolutionary 'intermediate state' of mitochondrial translation systems found in *Trichinella* species of parasitic nematodes: co-evolution of tRNA and EF-Tu. Nucleic Acids Research 34: 5291-5299.

Arlian, L.G. and Platts-Mills, T.A.E. (2001). The biology of dust mites and the remediation of mite allergens in allergic disease. Journal of Allergy and Clinical Immunology 107: S406-S413.

Arnason, U. and Johnsson, E. (1992). The complete mitochondrial DNA sequence of the harbor seal, *Phoca vitulina*. Journal of Molecular Evolution 34: 493-505.

Arnason, U., Gullberg, A., *et al.* (1993). The nucleotide sequence of the mitochondrial DNA molecule of the gray seal *Halichoerus grypus*, and a comparison with mitochondrial sequences of other true seals. Journal of Molecular Evolution 37: 323-330.

Ayroles, J.F., Carbone, M.A., *et al.* (2009). Systems genetics of complex traits in *Drosophila melanogaster*. Nature Genetics 41: 299-307.

Bariami, V., Jones, C.M., *et al.* (2012). Gene amplification, ABC transporters and cytochrome P450s: unraveling the molecular basis of pyrethroid resistance in the dengue vector, *Aedes aegypti*. PLoS Neglected Tropical Diseases 6: e1692.

Barker, F.K. and Lutzoni, F.M. (2002). The utility of the incongruence length difference test. Systematic Biology 51: 625-637.

Barker, S.C. and Murrell, A. (2002). Phylogeny, evolution and historical zoogeography of ticks: a review of recent progress. Experimental and Applied Acarology 28: 55-68.

Barthelme, D., Dinkelaker, S., *et al.* (2011). Ribosome recycling depends on a mechanistic link between the FeS cluster domain and a conformational switch of the twin-ATPase ABCE1. Proceedings of the National Academy of Sciences of the United States of America 108: 3228-3233.

Bass, C., Puinean, A.M., *et al.* (2011). Mutation of a nicotinic acetylcholine receptor beta subunit is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*. BMC Neuroscience 12: 51.

Bass, B., Hundley, H., *et al.* (2012). The difficult calls in RNA editing. Nature Biotechnology 30: 1207-1209.

Baum, J.A., Bogaert, T., *et al.* (2007). Control of coleopteran insect pests through RNA interference. Nature Biotechnology 25: 1322-1326.

Baxter, S.W., Badenes-Perez, F.R., *et al.* (2011). Parallel evolution of *Bacillus thuringiensis* toxin resistance in Lepidoptera. Genetics 189: 675-U814.

Baycin-Hizal, D., Tian, Y., *et al.* (2011). GlycoFly: a database of *Drosophila* N-linked glycoproteins identified using SPEG-MS techniques. Journal of Proteome Research 10: 2777-2784.

Baynes, R. (2009). Ectoparasiticides. *In* Veterinary pharmacology and therapeutics, J. E. Riviere and M. G. Papich, *Eds*. Wiley-Blackwell, Ames, Iowa**:** 1181-1201.

Beers, M.F., Hawkins, A., *et al.* (2011). A novel conserved targeting motif found in ABCA transporters mediates trafficking to early post-Golgi compartments. Journal of Lipid Research 52: 1471-1482.

Belles, X. (2010). Beyond *Drosophila*: RNAi In vivo and functional genomics in insects. Annual Review of Entomology 55: 111-128.

Bendtsen, J.D., Nielsen, H., *et al.* (2004). Improved prediction of signal peptides: SignalP 3.0. Journal of Molecular Biology 340: 783-795.

Berard, S., Bergeron, A., *et al.* (2007). Perfect sorting by reversals is not always difficult. Ieee/Acm Transactions on Computational Biology and Bioinformatics 4: 4-16.

Berenbaum, M.R., Cohen, M.B., *et al.* (1992). Cytochrome-P450 monooxygenase genes in oligophagous Lepidoptera. ACS Symposium Series 505: 114-124.

Berenbaum, M.R., Favret, C., *et al.* (1996). On defining key innovations in an adaptive radiation - cytochrome P450s and Papilionidae. American Naturalist 148: S139-155.

Berenbaum, M.R. (2002). Postgenomic chemical ecology: From genetic code to ecological interactions. Journal of Chemical Ecology 28: 873-896.

Bernays, E. and Graham, M. (1988). On the evolution of host specificity in phytophagous arthropods. Ecology 69: 886-892.

Bernays, E.A. and Chapman, R.F. (1994). Host-plant selection by phytophagous insects. Chapman & Hall, New York**:** 312 p.

Bernsel, A., Viklund, H., *et al.* (2008). Prediction of membrane-protein topology from first principles. Proceedings of the National Academy of Sciences of the United States of America 105: 7177-7181.

Bernt, M., Merkle, D., *et al.* (2007). CREx: inferring genomic rearrangements based on common intervals. Bioinformatics 23: 2957-2958.

Bernt, M., Braband, A., *et al.* (2013). Genetic aspects of mitochondrial genome evolution. Molecular Phylogenetics and Evolution: in press, doi: 10.1016/j.ympev.2012.1010.1020.

Berry, R.E., Yu, S.J., *et al.* (1980). Influence of host plants on insecticide metabolism and management of variegated cutworm (Lepidoptera, Noctuidae). Journal of Economic Entomology 73: 771-774.

Berthier, F., Renaud, M., *et al.* (1986). RNA mapping on Drosophila mitochondrial DNA - precursors and template strands. Nucleic Acids Research 14: 4519-4533.

Bi, J.L. and Felton, G.W. (1995). Foliar oxidative stress and insect herbivory - primary compounds, secondary metabolites, and reactive oxygen species as components of induced resistance. Journal of Chemical Ecology 21: 1511-1530.

Bischoff, E. (1988). Méthodes actuelles de quantification des acariens dans l'habitat. Revue Francaise d'Allergologie et d'Immunologie Clinique 28: 115-122.

Bissinger, B.W., Donohue, K.V., *et al.* (2011). Synganglion transcriptome and developmental global gene expression in adult females of the American dog tick, *Dermacentor variabilis* (Acari: Ixodidae). Insect Molecular Biology 20: 465-491.

Bitume, E.V., Bonte, D., *et al.* (2011). Heritability and artificial selection on ambulatory dispersal distance in *Tetranychus urticae*: effects of density and maternal effects. PLoS One 6: e26927.

Black IV, W.C. and Vontas, J.G. (2007). Affordable assays for genotyping single nucleotide polymorphisms in insects. Insect Molecular Biology 16: 377-387.

Black, W.C. and Roehrdanz, R.L. (1998). Mitochondrial gene order is not conserved in arthropods: Prostriate and metastriate tick mitochondrial genomes. Molecular Biology and Evolution 15: 1772-1785.

Black, W.C. and Bernhardt, S.A. (2009). Abundant nuclear copies of mitochondrial origin (NUMTs) in the *Aedes aegypti* genome. Insect Molecular Biology 18: 705-713.

Bleeker, P.M., Mirabella, R., *et al.* (2012). Improved herbivore resistance in cultivated tomato with the sesquiterpene biosynthetic pathway from a wild relative. Proceedings of the National Academy of Sciences of the United States of America 109: 20124-20129.

Bloomquist, J.R. (2003). Chloride channels as tools for developing selective insecticides. Archives of Insect Biochemistry and Physiology 54: 145-156.

Boerjan, B., Cardoen, D., *et al.* (2012). Insect omics research coming of age. Canadian Journal of Zoology-Revue Canadienne De Zoologie 90: 440-455.

Bonasio, R., Zhang, G., *et al.* (2010). Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. Science 329: 1068-1071.

Boore, J.L., Lavrov, D.V., *et al.* (1998). Gene translocation links insects and crustaceans. Nature 392: 667-668.

Boore, J.L. (1999). Animal mitochondrial genomes. Nucleic Acids Research 27: 1767-1780.

Boore, J.L. (2000). The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. *In* Comparative genomics, computational biology series, D. Sankoff and J. Nadeau, *Eds*. Kluwer Academic Publishers, Dordrecht**:** 133-147.

Boore, J.L., Macey, J.R., *et al.* (2005). Sequencing and comparing whole mitochondrial genomes of animals. *In* Molecular Evolution: Producing the Biochemical Data, Part B. Elsevier Academic Press Inc, San Diego**:** 311-348.

Boore, J.L. (2006a). The use of genome-level characters for phylogenetic reconstruction. Trends in Ecology & Evolution 21: 439-446.

Boore, J.L. (2006b). The complete sequence of the mitochondrial genome of *Nautilus macromphalus* (Mollusca : Cephalopoda). BMC Genomics 7: 182.

Borst, P., Evers, R., *et al.* (2000). A family of drug transporters: The multidrug resistance-associated proteins. Journal of the National Cancer Institute 92: 1295-1302.

Borst, P. and Elferink, R.O. (2002). Mammalian ABC transporters in health and disease. Annual Review of Biochemistry 71: 537-592.

Borycz, J., Borycz, J.A., *et al.* (2008). *Drosophila* ABC transporter mutants white, brown and scarlet have altered contents and distribution of biogenic amines in the brain. Journal of Experimental Biology 211: 3454-3466.

Boyce, T.M., Zwick, M.E., *et al.* (1989). Mitochondrial DNA in the bark weevils - Size, structure and heteroplasmy. Genetics 123: 825-836.

Brattsten, L., Wilkinson, C., *et al.* (1977). Herbivore-plant interactions: mixed-function oxidases and secondary plant substances. Science 196: 1349-1352.

Breton, S., Beaupré, H.D., *et al.* (2007). The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? Trends in Genetics 23: 465-474.

Breustedt, D.A., Schonfeld, D.L., *et al.* (2006). Comparative ligand-binding analysis of ten human lipocalins. Biochimica Et Biophysica Acta-Proteins and Proteomics 1764: 161-173.

Broehan, G., Kroeger, T., *et al.* (2013). Functional analysis of the ATP-binding cassette (ABC) transporter gene family of *Tribolium castaneum*. BMC Genomics 14: 6.

Brown, T.A. (2002). Genomes, 2nd edition. Wiley-Liss, Oxford**:** 608 p.

Brown, T.A., Cecconi, C., *et al.* (2005). Replication of mitochondrial DNA occurs by strand displacement with alternative light-strand origins, not via a strand-coupled mechanism. Genes & Development 19: 2466-2476.

Bryon, A., Van Leeuwen, T., *et al.* (2013). A focus on winter survival: Diapause and its gene regulation in the two-spotted spider mite. 18th National Symposium on Applied Biological Sciences, February 8th, Gent, Belgium.

Buckingham, S.D., Biggin, P.C., *et al.* (2005). Insect GABA receptors: Splicing, editing, and targeting by antiparasitics and insecticides. Molecular Pharmacology 68: 942-951.

Burger, T.D., Shao, R., *et al.* (2012). Phylogenetic analysis of ticks (Acari: Ixodida) using mitochondrial genomes and nuclear rRNA genes indicates that the genus *Amblyomma* is polyphyletic. Molecular Phylogenetics and Evolution 64: 45-55.

Burgess, S.T.G., Nisbet, A., *et al.* (2011). Generation, analysis and functional annotation of expressed sequence tags from the ectoparasitic mite *Psoroptes ovis*. Parasites and Vectors 4: 145.

Burgess, S.T.G., Downing, A., *et al.* (2012). Development of a cDNA microarray for the measurement of gene expression in the sheep scab mite *Psoroptes ovis*. Parasites & Vectors 5.

Buss, D.S. and Callaghan, A. (2008). Interaction of pesticides with p-glycoprotein and other ABC proteins: A survey of the possible importance to insecticide, herbicide and fungicide resistance. Pesticide Biochemistry and Physiology 90: 141-153.

Cabrera, A.R., Donohue, K.V., *et al.* (2011). New approach for the study of mite reproduction: The first transcriptome analysis of a mite, *Phytoseiulus persimilis* (Acari: Phytoseiidae). Journal of Insect Physiology 57: 52-61.

***Caenorhabditis elegans* Sequencing Consortium (1998)**. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. Science 282: 2012-2018.

Cameron, S.L., Miller, K.B., *et al.* (2004). Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea sensu lato (Arthropoda). Cladistics 20: 534-557.

Cameron, S.L., Beckenbach, A.T., *et al.* (2006). Evidence from mitochondrial genomics on interordinal relationhships in insects. Arthropod Systematics & Phylogeny 64: 27-34.

Cameron, S.L., Johnson, K.P., *et al.* (2007). The mitochondrial genome of the screamer louse Bothriometopus (Phthiraptera : Ischnocera): Effects of extensive gene rearrangements on the evolution of the genome. Journal of Molecular Evolution 65: 589-604.

Cameron, S.L. and Whiting, M.F. (2008). The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta*, (Insecta : Lepidoptera : Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. Gene 408: 112-123.

Campanella, J.J., Bitincka, L., *et al.* (2003). MatGAT: An application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinformatics 4: 29.

Campos, I., Geiger, J.A., *et al.* (2010). Genetic screen in Drosophila melanogaster uncovers a novel set of genes tequired for embryonic epithelial repair. Genetics 184: 129-U239.

Cao, Y., Adachi, J., *et al.* (1994). Phylogenetic-relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins - instability of a tree-based on a single-gene. Journal of Molecular Evolution 39: 519-527.

Carapelli, A., Comandi, S., *et al.* (2008). The complete mitochondrial genome of the antartic springtail *Cryptopygus antarcticus* (Hexapoda: Collembola). BMC Genomics 9: 315.

Carew, M., Schiffer, M., *et al.* (2009). Molecular markers indicate that the wheat curl mite, *Aceria tosichella* Keifer, may represent a species complex in Australia. Bulletin of Entomological Research 99: 479-486.

Carolan, J.C., Fitzroy, C.I.J., *et al.* (2009). The secreted salivary proteome of the pea aphid *Acyrthosiphon pisum* characterised by mass spectrometry. Proteomics 9: 2457-2467.

Carriere, Y. (2003). Haplodiploidy, sex, and the evolution of pesticide resistance. Journal of Economic Entomology 96: 1626-1640.

Castle, S.J., Prabhaker, N., *et al.* (2009). Host plant influence on susceptibility of *Bemisia tabaci* (Hemiptera: Aleyrodidae) to insecticides. Bulletin of Entomological Research 99: 263-273.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution 17: 540-552.

Castro, L.R. and Dowton, M. (2007). Mitochondrial genomes in the Hymenoptera and their utility as phylogenetic markers. Systematic Entomology 32: 60-69.

Celorio-Mancera, M.d.l.P., Heckel, D.G., *et al.* (2012). Transcriptional analysis of physiological pathways in a generalist herbivore: responses to different host plants and plant structures by the cotton bollworm, *Helicoverpa armigera*. Entomologia Experimentalis et Applicata 144: 123-133.

Chahine, S. and O'Donnell, M.J. (2009). Physiological and molecular characterization of methotrexate transport by Malpighian tubules of adult *Drosophila melanogaster*. Journal of Insect Physiology 55: 927-935.

Chahine, S. and O'Donnell, M.J. (2011). Interactions between detoxification mechanisms and excretion in Malpighian tubules of *Drosophila melanogaster*. Journal of Experimental Biology 214: 462-468.

Chamaon, K., Smalla, K.H., *et al.* (2002). Nicotinic acetylcholine receptors of *Drosophila*: three subunits encoded by genomically linked genes can co-assemble into the same receptor complex. Journal of Neurochemistry 80: 149-157.

Charron, J.-B.F., Ouellet, F., *et al.* (2008). The plant apolipoprotein D ortholog protects *Arabidopsis* against oxidative stress. BMC Plant Biology 8: 86.

Chen, H., Wilkerson, C.G., *et al.* (2005). Jasmonate-inducible plant enzymes degrade essential amino acids in the herbivore midgut. Proceedings of the National Academy of Sciences of the United States of America 102: 19237-19242.

Chen, X., Yu, Z., *et al.* (2012). Life cycle of *Haemaphysalis doenitzi* (Acari: Ixodidae) under laboratory conditions and its phylogeny based on mitochondrial 16S rDNA. Experimental and Applied Acarology 56: 143-150.

Chilana, P., Sharma, A., *et al.* (2012). Insect genomic resources: status, availability and future. Current Science 102: 571-580.

Chmelar, J., Calvo, E., *et al.* (2012). Tick salivary secretion as a source of antihemostatics. Journal of Proteomics 75: 3842-3854.

Chudzinski-Tavassi, A.M., Carrijo-Carvalho, L.C., *et al.* (2010). A lipocalin sequence signature modulates cell survival. FEBS Letters 584: 2896-2900.

Cilia, M., Tamborindeguy, C., *et al.* (2011). Tangible benefits of the aphid *Acyrthosiphon pisum* genome sequencing for aphid proteomics: Enhancements in protein identification and data validation for homology-based proteomics. Journal of Insect Physiology 57: 179-190.

Clark, A.G., Eisen, M.B., *et al.* (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450: 203-218.

Claudianos, C., Ranson, H., *et al.* (2006). A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. Insect Molecular Biology 15: 615-636.

Clayton, D.A. (1991). Replication and transcription of vertebrate mitochondrial DNA. Annual Review of Cell Biology 7: 453-478.

Colbourne, J.K., Pfrender, M.E., *et al.* (2011). The ecoresponsive genome of *Daphnia pulex*. Science 331: 555-561.

Cole, C., Barber, J.D., *et al.* (2008). The Jpred 3 secondary structure prediction server. Nucleic Acids Research 36: W197-W201.

Cook, D.C., Thomas, M.B., *et al.* (2007). Predicting the economic impact of an invasive species on an ecosystem service. Ecological Applications 17: 1832-1840.

Cornell, H.V. and Hawkins, B.A. (2003). Herbivore responses to plant secondary compounds: a test of phytochemical coevolution theory. American Naturalist 161: 507-522.

Cornman, S.R., Schatz, M.C., *et al.* (2010). Genomic survey of the ectoparasitic mite *Varroa destructor*, a major pest of the honey bee *Apis mellifera*. BMC Genomics 11: 602.

Croft, B.A. and Strickler, K. (1983). Natural enemy resistance to pesticides: documentation, characterization, theory and application. *In* Pest resistance to pesticides, G. P. Georghiou and T. Saito, *Eds*. Plenum, New York, USA**:** 669-702.

Cui, P., Ji, R., *et al.* (2007). A complete mitochondrial genome sequence of the wild two-humped camel (*Camelus bactrianus ferus*): an evolutionary history of camelidae. BMC Genomics 8: 241.

Cully, D.F., Vassilatis, D.K., *et al.* (1994). Cloning of an avermectin sensitive glutamate-gated chloride channel from *Caenorhabditis elegans*. Nature 371: 707-711.

Cunningham, C.W. (1997). Can three incongruence tests predict when data should be combined? Molecular Biology and Evolution 14: 733-740.

Cunningham, C.W., Zhu, H., *et al.* (1998). Best-fit maximum-likelihood models for phylogenetic inference: Empirical tests with known phylogenies. Evolution 52: 978-987.

Curole, J.P. and Kocher, T.D. (1999). Mitogenomics: digging deeper with complete mitochondrial genomes. Trends in Ecology & Evolution 14: 394-398.

Dabert, M. (2006). DNA markers in the phylogenetics of the Acari. Biological Letters 43: 97-107.

Dabert, M., Witalinski, W., *et al.* (2010). Molecular phylogeny of acariform mites (Acari, Arachnida): Strong conflict between phylogenetic signal and long-branch attraction artifacts. Molecular Phylogenetics and Evolution 56: 222-241.

Daborn, P.J., Yen, J.L., *et al.* (2002). A single P450 allele associated with insecticide resistance in *Drosophila*. Science 297: 2253-2256.

Dale, R.P., Jones, A.K., *et al.* (2010). Identification of ion channel genes in the *Acyrthosiphon pisum* genome. Insect Molecular Biology 19: 141-153.

Dasmahapatra, K.K., Walters, J.R., *et al.* (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487: 94-98.

David, J.P., Strode, C., *et al.* (2005). The *Anopheles gambiae* detoxification chip: A highly specific microarray to study metabolic-based insecticide resistance in malaria vectors. Proceedings of the National Academy of Sciences of the United States of America 102: 4080-4084.

Davila, S., Pinero, D., *et al.* (2005). The mitochondrial genome sequence of the scorpion *Centruroides limpidus* (Karsch 1879) (Chelicerata; Arachnida). Gene 360: 92-102.

De Jong, G. (1991). Long-term health-effects of aldrin and dieldrin - a study of exposure, health-effects and mortality of workers engaged in the manufacture and formulation of the insecticides aldrin and dieldrin. Toxicology Letters: 1-206.

de la Fuente, J. and Kocan, K.M. (2006). Strategies for development of vaccines for control of ixodid tick species. Parasite Immunology 28: 275-283.

de la Fuente, J., Almazan, C., *et al.* (2006). Autocidal control of ticks by silencing of a single gene by RNA interference. Biochemical and Biophysical Research Communications 344: 332-338.

de la Fuente, J., Kocan, K.M., *et al.* (2007a). RNA interference for the study and genetic manipulation of ticks. Trends in Parasitology 23: 427-433.

de la Fuente, J., Almazán, C., *et al.* (2007b). A ten-year review of commercial vaccine performance for control of tick infestations on cattle. Animal Health Research Reviews 8: 23-28.

De Lillo, E. and Monfreda, R. (2004). 'Salivary secretions' of eriophyoids (Acari : Eriophyoidea): first results of an experimental model. Experimental and Applied Acarology 34: 291-306.

De Rijk, P., Robbrecht, E., *et al.* (1999). Database on the structure of large subunit RNA. Nucleic Acids Research 27: 174-178.

De Rijk, P., Wuyts, J., *et al.* (2003). RnaViz 2: an improved representation of RNA secondary structure. Bioinformatics 19: 299-300.

Dean, F. B., J. R. Nelson, *et al.* (2001). Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. Genome Research 11: 1095-1099.

Dean, M., Rzhetsky, A., *et al.* (2001). The human ATP-binding cassette (ABC) transporter superfamily. Genome Research 11: 1156-1166.

Dean, M. and Annilo, T. (2005). Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. *In* Annual Review of Genomics and Human Genetics. Vol. 6, D. S. Page and A. Chakravarti, *Eds*. Palo Alto, California**:** 123-142.

Deeley, R.G., Westlake, C., *et al.* (2006). Transmembrane transport of endo- and xenobiotics by mammalian ATP-binding cassette multidrug resistance proteins. Physiological Reviews 86: 849-899.

Dekeyser, M.A. (2005). Acaricide mode of action. Pest Management Science 61: 103-110.

Demaeght, P., Dermauw, W., *et al.* (2013). Molecular analysis of resistance to acaricidal spirocyclic tetronic acids in *Tetranychus urticae*: CYP392E10 metabolizes spirodiclofen, but not its corresponding enol. Insect Biochemistry and Molecular Biology, in press.

Denholm, I., Cahill, M., *et al.* (1998). Challenges with managing insecticide resistance in agricultural pests, exemplified by the whitefly Bemisia tabaci. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 353: 1757-1767.

Dennehy, T., Nyrop, J., *et al.* (1990). Characterization and exploitation of instability of spider mite. *In* Managing resistance to agrochemicals: from fundamental research to practical strategies. , G. MB, L. H and M. WK, *Eds*. American Chemical Society, Washington**:** 77–91.

Dent, J.A. (2006). Evidence for a diverse cys-loop ligand-gated ion channel superfamily in early bilateria. Journal of Molecular Evolution 62: 523-535.

Desalle, R. and Rosenfeld, J. (2012). Phylogenomics: a primer. Garland Science**:** 352 p.

Despres, L., David, J., *et al.* (2007). The evolutionary ecology of insect resistance to plant chemicals. Trends in Ecology & Evolution 22: 298-307.

Dolphin, K., Belshaw, R., *et al.* (2000). Noise and incongruence: Interpreting results of the incongruence length difference test. Molecular Phylogenetics and Evolution 17: 401-406.

Domes, K., R. A. Norton, *et al.* (2007). Reevolution of sexuality breaks Dollo's law. Proceedings of the National Academy of Sciences 104: 7139-7144.

Domes, K., Maraun, M., *et al.* (2008). The complete mitochondrial genome of the sexual oribatid mite *Steganacarus magnus*: genome rearrangements and loss of tRNAs. BMC Genomics 9: 532.

Dowling, A.P.G. and Oconnor, B.M. (2010). Phylogenetic relationships within the suborder Dermanyssina (Acari: Parasitiformes) and a test of dermanyssoid monophyly. International Journal of Acarology 36: 299-312.

Dowton, M., Castro, L.R., *et al.* (2002). Mitochondrial gene rearrangements as phylogenetic characters in the invertebrates: the examination of genome 'morphology'. Invertebrate Systematics 16: 345-356.

Dowton, M. and Austin, A.D. (2002). Increased congruence does not necessarily indicate increased phylogenetic accuracy - The behavior of the incongruence length difference test in mixed-model analyses. Systematic Biology 51: 19-31.

Dunlop, J.A. and Alberti, G. (2008). The affinities of mites and ticks: a review. Journal of Zoological Systematics and Evolutionary Research 46: 1-18.

Dunlop, J.A. (2010). Geological history and phylogeny of Chelicerata. Arthropod Structure & Development 39: 124-142.

Dutta, S. (2008). Genetic regulation of autophagic cell death in *Drosophila melanogaster*, University of Maryland, Department Molecular and Cell Biology. PhD**:** 130 p.

Dyck, V.A., Hendrichs, J., ***et al.*** **(2005)**. Sterile Insect Technique: principles and practice in area-wide integrated pest management. Springer, London, UK**:** 787 p.

Edgar, R.C. (2004a). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 1-19.

Edgar, R.C. (2004b). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-1797.

Edwards, D.D., Jackson, L.E., ***et al.*** **(2011)**. Mitochondrial genome sequence of *Unionicola parkeri* (Acari: Trombidiformes: Unionicolidae): molecular synapomorphies between closely-related Unionicola gill mites. Experimental and Applied Acarology 54: 105-117.

Eisen, J.A. and Fraser, C.M. (2003). Phylogenomics: Intersection of evolution and genomics. Science 300: 1706-1707.

Eleftherianos, I., Won, S., ***et al.*** **(2011)**. ATP-sensitive potassium channel (K-ATP)-dependent regulation of cardiotropic viral infections. Proceedings of the National Academy of Sciences of the United States of America 108: 12024-12029.

Elliott, A.M. and Ai-Hajj, M.A. (2009). ABCB8 Mediates doxorubicin resistance in melanoma cells by protecting the mitochondrial genome. Molecular Cancer Research 7: 79-87.

Emanuelsson, O., Brunak, S., ***et al.*** **(2007)**. Locating proteins in the cell using TargetP, SignalP and related tools. Nature Protocols 2: 953-971.

Embers, M.E. and Narasimhan, S. (2013). Vaccination against Lyme disease: past, present and future. Frontiers in Cellular and Infection Microbiology 3: in press, doi: 10.3389/fcimb.2013.00006.

Emous, R.v., Fiks-van Niekerk, T., ***et al.*** **(2005)**. Bloedluizen (vogelmijten) op papier en in de praktijk - red mites in theory and practice. Animal Sciences Group/Praktijkonderzoek, Lelystad, The Netherlands**:** 44 p.

Ernsting, B.R., Edwards, D.D., ***et al.*** **(2009)**. Mitochondrial genome sequence of *Unionicola foili* (Acari: Unionicolidae): a unique gene order with implications for phylogenetic inference. Experimental and Applied Acarology 49: 305-316.

Fahrein, K., Talarico, G., ***et al.*** **(2007)**. The complete mitochondrial genome of *Pseudocellus pearsei* (Chelicerata : Ricinulei) and a comparison of mitochondrial gene rearrangements in Arachnida. BMC Genomics 8: 386.

Farris, J.S., Kallersjo, M., ***et al.*** **(1995)**. Constructing a significance test for incongruence. Systematic Biology 44: 570-572.

Feijao, P.C., Neiva, L.S., ***et al.*** **(2006)**. AMiGA: the arthropodan mitochondrial genomes accessible database. Bioinformatics 22: 902-903.

Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positvely misleading. Systematic Zoology 27: 401-410.

Ferraroni, M., Seifert, J., ***et al.*** **(2005)**. Crystal structure of the hydroxyquinol 1,2-dioxygenase from *Nocardioides simplex* 3E, a key enzyme involved in polychlorinated aromatics biodegradation. Journal of Biological Chemistry 280: 21144-21154.

Feyereisen, R. (2011). Arthropod CYPomes illustrate the tempo and mode in P450 evolution. Biochimica Et Biophysica Acta 1814: 19-28.

Feyereisen, R. (2012). Insect CYP Genes and P450 Enzymes. *In* Insect molecular biology and biochemistry, L. I. Gilbert, *Ed*. Academic Press, Elsevier, London, UK**:** 236-316.

ffrench-Constant, R.H., Daborn, P.J., ***et al.*** **(2004)**. The genetics and genomics of insecticide resistance. Trends in Genetics 20: 163-170.

Fiers, W., Contreras, R., ***et al.*** **(1976)**. Complete nucleotide sequence of bacteriophage MS2-RNA - Primary and secondrary structure of replicase gene. Nature 260: 500-507.

Flower, D.R., North, A.C.T., *et al.* (2000). The lipocalin protein family: structural and sequence overview. Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology 1482: 9-24.

Fortin, P.D., Lo, A.T.-F., *et al.* (2005). Evolutionarily divergent extradiol dioxygenases possess higher specificities for polychlorinated biphenyl metabolites. Journal of Bacteriology 187: 415-421.

Francino, M.P. and Ochman, H. (1997). Strand asymmetries in DNA evolution. Trends in Genetics 13: 240-245.

Friedrich, M. and Tautz, D. (1995). Ribosomal dna phylogeny of the major extant arthropod classes and the evolution of myriapods. Nature 376: 165-167.

Fry, J.D. (1989). Evolutionary adaptation to host plants in a laboratory population of the phytophagous mite *Tetranychus urticae* Koch. Oecologia 81: 559-565.

Fujita, M.K., Boore, J.L., *et al.* (2007). Multiple origins and rapid evolution of duplicated mitochondrial genes in parthenogenetic geckos (*Heteronotia binoei*; Squamata, Gekkonidae). Molecular Biology and Evolution 24: 2775-2786.

Fukuda, Y. and Schuetz, J.D. (2012). ABC transporters and their role in nucleoside and nucleotide drug resistance. Biochemical Pharmacology 83: 1073-1083.

Futuyma, D.J. and Peterson, S.C. (1985). Genetic variation in the use of resources by insects. Annual Review of Entomology 30: 217-238.

Futuyma, D.J. and Agrawal, A.A. (2009). Macroevolution and the biological diversity of plants and herbivores. Proceedings of the National Academy of Sciences of the United States of America 106: 18054-18061.

Gaertner, L.S., Murray, C.L., *et al.* (1998). Transepithelial transport of nicotine and vinblastine in isolated malpighian tubules of the tobacco hornworm (*Manduca sexta*) suggests a P-glycoprotein-like mechanism. Journal of Experimental Biology 201: 2637-2645.

Galzi, J.L., Bertrand, D., *et al.* (1991). Functional significance of aromatic-amino-acids from 3 peptide loops of the alfa-7 neuronal nicotinic receptor-site investigated by site-directed mutagenesis. FEBS Letters 294: 198-202.

Ganfornina, M.D., Gutierrez, G., *et al.* (2000). A phylogenetic analysis of the lipocalin protein family. Molecular Biology and Evolution 17: 114-126.

Ganfornina, M.D., Kayser, H., *et al.* (2006). Lipocalins in Arthropoda: diversification and functional Explorations. *In* Lipocalins, B. Åkerström, N. Borregaard, D. R. Flower and J. P. Salier, *Eds*. Landes Bioscience, Austin, Texas, USA**:** 49-74.

Ganfornina, M.D., Do Carmo, S., *et al.* (2008). Apolipoprotein D is involved in the mechanisms regulating protection from oxidative stress. Aging Cell 7: 506-515.

Gangishetti, U., Breitenbach, S., *et al.* (2009). Effects of benzoylphenylurea on chitin synthesis and orientation in the cuticle of the *Drosophila* larva. European Journal of Cell Biology 88: 167-180.

Garbian, Y., Maori, E., *et al.* (2012). Bidirectional transfer of RNAi between honey bee and *Varroa destructor*: *Varroa* gene silencing reduces *Varroa* population. PLoS Pathogens 8: e1003035.

George, J.E., Pound, J.M., *et al.* (2004). Chemical control of ticks on cattle and the resistance of these parasites to acaricides. Parasitology 129: S353-S366.

Georghiou, G. (1969). Genetics resistance to insecticides in houseflies and mosquitoes. Experimental Parasitology 26: 224-255.

Georghiou, G.P. and Taylor, C.E. (1977). Genetic and biological influences in the evolution of insecticide resistance. Journal of Economic Entomology 70: 319-323.

Georgiev, P.G., Wolstenholme, A.J., *et al.* (2002). Differential responses to avermectins in ort mutants of *Drosophila melanogaster*. Pesticide Biochemistry and Physiology 72: 65-71.

Geraci, N.S., Johnston, J.S., *et al.* (2007). Variation in genome size of argasid and ixodid ticks. Insect Biochemistry and Molecular Biology 37: 399-408.

Gerson, U., Smiley, R.L., *et al.* (2003). Mites (Acari) for pest control. Blackwell Science, Oxford**:** 539 p.

Giribet, G. and Edgecombe, G.D. (2012). Reevaluating the arthropod tree of life. Annual Review of Entomology 57: 167-186.

Gissi, C. and Pesole, G. (2003). Transcript mapping and genome annotation of ascidian mtDNA using EST data. Genome Research 13: 2203-2212.

Gissi, C., Iannelli, F.*, et al.* (2008). Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity 101: 301-320.

Glastad, K.M., Hunt, B.G.*, et al.* (2011). DNA methylation in insects: on the brink of the epigenomic era. Insect Molecular Biology 20: 553-565.

Goldstein, D.B. and Schlotterer, C. (1999). Microsatellites: evolution and applications. Oxford University Press, USA**:** 368 p.

Gonzalez-Cabo, P., Bolinches-Amoros, A.*, et al.* (2011). Disruption of the ATP-binding cassette B7 (ABTM-1/ABCB7) induces oxidative stress and premature cell death in *Caenorhabditis elegans*. Journal of Biological Chemistry 286: 21304-21314.

Goodrich-Blair, H. and Clarke, D.J. (2007). Mutualism and pathogenesis in *Xenorhabdus* and *Photorhabdus*: two roads to the same destination. Molecular Microbiology 64: 260-268.

Gordon, H.T. (1961). Nutritional factors in insect resistance to chemicals. Annual Review of Entomology 6: 27-54.

Gorrochotegui-Escalante, N. and Black, W.C. (2003). Amplifying whole insect genomes with multiple displacement amplification. Insect Molecular Biology 12: 195-200.

Gotoh, T., Bruin, J.*, et al.* (1993). Host race formation in *Tetranychus urticae* - Genetic differentiation, host-plant preference, and mate chocie in a tomato and a cucumber strain. Entomologia Experimentalis et Applicata 68: 171-178.

Gough, N. (1990). Evaluation of miticides for the control of 2-spotted mite *Tetranychus urticae* KOCH on field roses in Southern Queensland. Crop Protection 9: 119-127.

Gould, F. (1979). Rapid host range evolution in a population of the phytophagous mite *Tetranychus urticae* Koch. Evolution 33: 791-802.

Gould, F., Carroll, C.R.*, et al.* (1982). Cross-resistance to pesticides and plant defenses: a study of the two-spotted spider-mite. Entomologia Experimentalis Et Applicata 31: 175-180.

Govind, G., Mittapalli, O.*, et al.* (2010). Unbiased transcriptional comparisons of generalist and specialist herbivores feeding on progressively defenseless *Nicotiana attenuata* plants. PLoS One 5: e8735.

Graf, G.A., Yu, L.Q.*, et al.* (2003). ABCG5 and ABCG8 are obligate heterodimers for protein trafficking and biliary cholesterol excretion. Journal of Biological Chemistry 278: 48275-48282.

Grbić, M., Van Leeuwen, T.*, et al.* (2011). The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. Nature 479: 487-492.

Green, M.M. (1996). The "genesis of the white-eyed mutant" in *Drosophila melanogaster*: A reappraisal. Genetics 142: 329-331.

Gu, H., Smith, Z.D.*, et al.* (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nature Protocols 6: 468-481.

Guerrero, F. D., V. M. Nene*, et al.* (2006). Sequencing a new target genome: The *Boophilus microplus* (Acari: Ixodidae) genome project. Journal of Medical Entomology 43: 9-16.

Guerrero, F.D., Moolhuijzen, P.*, et al.* (2010). Reassociation kinetics-based approach for partial genome sequencing of the cattle tick, *Rhipicephalus* (*Boophilus*) *microplus*. BMC Genomics 11: 374.

Guerrero, F.D., Miller, R.J.*, et al.* (2012). Cattle tick vaccines: Many candidate antigens, but will a commercially viable product emerge? International Journal for Parasitology 42: 421-427.

Hackett, K.J. (2011). The 5,000 insect and other arthropod genome initiative. 5th Arthropod Genomics Consortium, June 9-12th, Kansas, USA.

Hall, T. (1999). BioEdit: a user-friendly biological sequence alignment-editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41: 95-98.

Hanrahan, S.J. and Johnston, J.S. (2011). New genome size estimates of 134 species of arthropods. Chromosome Research 19: 809-823.

Hansen, K.K., Kristensen, M.*, et al.* (2005). Correlation of a resistance-associated Rdl mutation in the German cockroach, *Blattella germanica* (L), with persistent dieldrin resistance in two Danish field populations. Pest Management Science 61: 749-753.

Hart, B. (1990). Ecology and biology of allergenic mites. *In* Mites and allergic disease, Guerin, *Ed*. Groeninge Press, Belgium**:** 135-152.

Harwood, C.S. and Parales, R.E. (1996). The beta-ketoadipate pathway and the biology of self-identity. Annual Review of Microbiology 50: 553-590.

Hase, A. (1929). Zur pathologisch-parasitologischen und epidemiologisch-hygienischen Bedeutung der Milben, ins besondere der Tyroglyphinae (Käsemilben), sowie über den sogenannten „Milbenkäse “. Parasitology Research 1: 765-821.

Hassanin, A., Leger, N., *et al.* (2005). Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. Systematic Biology 54: 277-298.

Hassanin, A. (2006). Phylogeny of Arthropoda inferred from mitochondrial sequences: Strategies for limiting the misleading effects of multiple changes in pattern and rates of substitution. Molecular Phylogenetics and Evolution 38: 100-116.

Haupt, J. and Coineau, Y. (1999). Ultrastructure and functional morphology of a nematalycid mite (Acari : Actinotrichida : Endostigmata : Nematalycidae): adaptations to mesopsammal life. Acta Zoologica 80: 97-111.

Hazkani-Covo, E., Zeller, R.M., *et al.* (2010). Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genetics 6: e1000834.

He, Y., Jones, J., *et al.* (2005). The mitochondrial genome of *Xiphinema americanum* sensu stricto (Nematoda : Enoplea): Considerable economization in the length and structural features of encoded genes. Journal of Molecular Evolution 61: 819-833.

He, L., Gao, X., *et al.* (2009). Genetic analysis of abamectin resistance in *Tetranychus cinnabarinus*. Pesticide Biochemistry and Physiology 95: 147-151.

Heckel, D.G. (2003). Genomics in pure and applied entomology. Annual Review of Entomology 48: 235-260.

Heckel, D.G. (2012). Learning the ABCs of Bt: ABC transporters and insect resistance to *Bacillus thuringiensis* provide clues to a crucial step in toxin mode of action. Pesticide Biochemistry and Physiology 104: 103-110.

Heethoff, M. and Koerner, L. (2007). Small but powerful: the oribatid mite *Archegozetes longisetosus* Aoki (Acari, Oribatida) produces disproportionately high forces. Journal of Experimental Biology 210: 3036-3042.

Helle, W. and Bolland, **R.H. (1967)**. Karyoptypes and sex-determination in spider mites (Tetranychidae). Genetica 28, 43-53.

Hengge, U.R., Currie, B.J., *et al.* (2006). Scabies: a ubiquitous neglected skin disease. Lancet Infectious Diseases 6: 769-779.

Hess, P.N. and De Moraes Russo, C.A. (2007). An empirical test of the midpoint rooting method. Biological Journal of the Linnean Society 92: 669-674.

Herz, J. and Strickland, D.K. (2001). LRP: a multifunctional scavenger and signaling receptor. Journal of Clinical Investigation 108: 779-784.

Hieter, P. and Boguski, M. (1997). Functional genomics: It's all how you read it. Science 278: 601-602.

Higgins, C.F. (1992). ABC transporters - from microorganisms to man. Annual Review of Cell Biology 8: 67-113.

Hildebrand, D.F., Rodriguez, J.G., *et al.* (1986). Peroxidative responses of leaves in 2 soybean genotypes injured by 2-spotted spider-mites (Acari, Tetranychidae). Journal of Economic Entomology 79: 1459-1465.

Hill, C.A. (2010). Genome analysis of major tick and mite vectors of human pathogens. Available online: http://www.genome.gov/Pages/Research/DER/PathogensandVectors/Tick_and_Mite_Genomes_Cluster_White_Paper_12Jan2011.pdf.

Hillen, W. and Berens, C. (1994). Mechanisms underlying expression of tn10 encoded tetracycline resistance. Annual Review of Microbiology 48: 345-369.

Hipp, A.L., Hall, J.C., *et al.* (2004). Congruence versus phylogenetic accuracy: Revisiting the incongruence length difference test. Systematic Biology 53: 81-89.

Hock, T., Cottrill, T., *et al.* (2000). The E23 early gene of *Drosophila* encodes an ecdysone-inducible ATP-binding cassette transporter capable of repressing ecdysone-mediated gene activation. Proceedings of the National Academy of Sciences of the United States of America 97: 9519-9524.

Hoelzel, A.R., Hancock, J.M., *et al.* (1993). Generations of VNTRS and heteroplasmy by sequence turnover in the mitochondrial control region of 2 elephant seal species. Journal of Molecular Evolution 37: 190-197.

Hollenstein, K., Dawson, R.J.P., *et al.* (2007). Structure and mechanism of ABC transporter proteins. Current Opinion in Structural Biology 17: 412-418.

Hong, S.-T., Bang, S., *et al.* (2006). Histamine and its receptors modulate temperature-preference behaviors in *Drosophila*. Journal of Neuroscience 26: 7245-7256.

Hope, M., Menzies, M., *et al.* (2010). Identification of a dieldrin resistance-associated mutation in *Rhipicephalus* (*Boophilus*) *microplus* (Acari: Ixodidae). Journal of Economic Entomology 103: 1355-1359.

Hopper-Borge, E., Xu, X., *et al.* (2009). Human Multidrug Resistance Protein 7 (ABCC10) is a resistance factor for nucleoside analogues and epothilone B. Cancer Research 69: 178-184.

Hosie, A.M., Aronstein, K., *et al.* (1997). Molecular biology of insect neuronal GABA receptors. Trends in Neurosciences 20: 578-583.

Hovius, J.W.R., Levi, M., *et al.* (2008). Salivating for knowledge: Potential pharmacological agents in tick saliva. PLoS Medicine 5: 202-208.

Howell, N. (1985). Anomalous electrophoretic mobility of mouse mtDNA restriction fragments. Plasmid 14: 93-96.

Hoy, M.A. (2009). The predatory mite *Metaseiulus occidentalis*: mitey small and mitey large genomes. BioEssays 31: 581-590.

Hoy, M.A., Yu, F., *et al.* (2013). Transcriptome sequencing and annotation of the predatory mite *Metaseiulus occidentalis* (Acari: Phytoseiidae): a cautionary tale about possible contamination by prey sequences. Experimental and Applied Acarology 59: 283-296.

Huang, H. and Haddad, G.G. (2007). *Drosophila* dMRP4 regulates responsiveness to O-2 deprivation and development under hypoxia. Physiological Genomics 29: 260-266.

Huang, D.W., Sherman, B.T., *et al.* (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 4: 44-57.

Huang, J.L.W. (2011). Developing and using expressed sequence tags to study the predatory mite *Phytoseiulus persimilis* Athias-Henriot (Paraistiformes, Mesostigmata, Phytoseiidae). Department of Entomology Manhattan, Kanas, Kansas State University. PhD**:** 154 p.

Huelsenbeck, J.P. and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754-755.

Hughes, T.R., Mao, M., *et al.* (2001). Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nature Biotechnology 19: 342-347.

Humair, P.-F., Douet, V., *et al.* (2007). Molecular identification of bloodmeal source in *Ixodes ricinus* ticks using 12S rDNA as a genetic marker. Journal of Medical Entomology 44: 869-880.

Hunter, W. and Chittarankan, K. (2008). Genome mapping and genomics in arthropods. Springer-Verlag, New York, USA**:** 122 p.

Ibrahim, M.M.S. (2009). Effect of the host plant on susceptibility of the two-spotted spider mite, *Tetranychus urticae* Koch, (Acari: Tetranychidae) to some acaricides. Mansoura University Journal of Agricultural Sciences 34: 10735 - 10744.

Ichikawa, Y., Bayeva, M., *et al*. (2012). Disruption of ATP-binding cassette B8 in mice leads to cardiomyopathy through a decrease in mitochondrial iron export**.** Proceedings of the National Academy of Sciences of the United States of America, 109:4152-4157.

Iovchev, M., Kodrov, P., *et al.* (2002). Altered drug resistance and recovery from paralysis in *Drosophila melanogaster* with a deficient histamine-gated chloride channel. Journal of Neurogenetics 16: 249-261.

Ishikawa, T. (1992). The ATP-dependent glutathione s-conjugate export pump. Trends in Biochemical Sciences 17: 463-469.

Ito, K. and Saito, Y. (2006). Effects of host-plant species on diapause induction of the Kanzawa spider mite, *Tetranychus kanzawai*. Entomologia Experimentalis et Applicata 121: 177-184.

Itoh, T.Q. and Matsumoto, A. (2011). Membrane-bound transporter controls the circadian transcription of clock genes in *Drosophila*. *Genes Cells*, 16:1159-1167.

James, D.G. and Price, T.S. (2002). Fecundity in twospotted spider mite (Acari : Tetranychidae) is increased by direct and systemic exposure to imidacloprid. Journal of Economic Entomology 95: 729-732.

Janson, C., Anto, J.*, et al.* (2001). The European Community Respiratory Health Survey: what are the main results so far? European Respiratory Journal 18: 598-611.

Jensen, M.L., Schousboe, A.*, et al.* (2005). Charge selectivity of the Cys-loop family of ligand-gated ion channels. Journal of Neurochemistry 92: 217-225.

Jeppson, L.R., Keifer, H.H.*, et al.* (1975). Mites injurious to economic plants. University of California Press, Berkeley, Los Angeles, USA**:** 614 p.

Jeschke, P. and Nauen, R. (2008). Neonicotinoids - from zero to hero in insecticide chemistry. Pest Management Science 64: 1084-1098.

Jeyaprakash, A. and Hoy, M.A. (2007). The mitochondrial genome of the predatory mite *Metaseiulus occidentalis* (Arthropoda : Chelicerata : Acari : Phytoseiidae) is unexpectedly large and contains several novel features. Gene 391: 264-274.

Jeyaprakash, A. and Hoy, M. (2009a). The nuclear genome of the phytoseiid *Metaseiulus occidentalis* (Acari: Phytoseiidae) is among the smallest known in arthropods. Experimental and Applied Acarology 47: 263-273.

Jeyaprakash, A. and Hoy, M. (2009b). First divergence time estimate of spiders, scorpions, mites and ticks (subphylum: Chelicerata) inferred from mitochondrial phylogeny. Experimental and Applied Acarology 47: 1-18.

Jobb, G., von Haeseler, A.*, et al.* (2004). TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evolutionary Biology 4: 18.

Jones, A.K. and Sattelle, D.B. (2006). The cys-loop ligand-gated ion channel superfamily of the honeybee, *Apis mellifera*. Invertebrate Neuroscience 6: 123-132.

Jones, A.K. and Sattelle, D.B. (2007). The cys-loop ligand-gated ion channel gene superfamily of the red flour beetle, *Tribolium castaneum*. BMC Genomics 8: 327.

Jones, A.K. and Sattelle, D.B. (2008). The cys-loop ligand-gated ion channel gene superfamily of the nematode, *Caenorhabditis elegans*. Invertebrate Neuroscience 8: 41-47.

Jones, A.K. and Sattelle, D.B. (2010). Diversity of insect nicotinic acetylcholine receptor subunits. *In* Insect nicotinic acetylcholine receptors, S. H. Thany, *Ed***:** 25-43.

Jones, A.K., Bera, A.N.*, et al.* (2010). The cys-loop ligand-gated ion channel gene superfamily of the parasitoid wasp, *Nasonia vitripennis*. Heredity 104: 247-259.

Jones, C.M., Toe, H.K.*, et al.* (2012). Additional selection for insecticide resistance in urban malaria vectors: DDT Resistance in *Anopheles arabiensis* from Bobo-Dioulasso, Burkina Faso. PLoS One 7.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics 13: 484-492.

Jongejan, F. and Uilenberg, G. (2004). The global importance of ticks. Parasitology 129: S3-S14.

Jordan, I.K., Kota, K.C.*, et al.* (2008). Evolutionary and functional divergence between the cystic fibrosis transmembrane conductance regulator and related ATP-binding cassette transporters. Proceedings of the National Academy of Sciences of the United States of America 105: 18865-18870.

Julenius, K., Molgaard, A.*, et al.* (2005). Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. Glycobiology 15: 153-164.

Kalpaklioglu, A.F., Ferizli, A.G.*, et al.* (1996). The effectiveness of benzyl benzoate and different chemicals as acaricides. Allergy 51: 164-170.

Kane, N.S., Hirschberg, B.*, et al.* (2000). Drug-resistant *Drosophila* indicate glutamate-gated chloride channels are targets for the antiparasitics nodulisporic acid and ivermectin. Proceedings of the National Academy of Sciences of the United States of America 97: 13949-13954.

Kant, M.R., Sabelis, M.W.*, et al.* (2008). Intraspecific variation in a generalist herbivore accounts for differential induction and impact of host plant defences. Proceedings of the Royal Society B-Biological Sciences Series B 275: 443-452.

Karim, S., Troiano, E.*, et al.* (2010). Functional genomics tool: gene silencing in *Ixodes scapularis* eggs and nymphs by electroporated dsRNA. BMC Biotechnology 10: 1.

Karim, S. and Adamson, S.W. (2012). RNA Interference in ticks: A functional genomics tool for the study of physiology. Advances in Insect Physiology 42: 119-154.

Kehoe, J., Buldakova, S.*, et al.* (2009). *Aplysia* cys-loop glutamate-gated chloride channels reveal convergent evolution of ligand specificity. Journal of Molecular Evolution 69: 125-141.

Kennedy, G.G. (1984). 2-tridecanone, tomatoes and *Heliothis zea* - potential incompatibility of plant antibiosis with insecticidal control. Entomologia Experimentalis Et Applicata 35: 305-311.

Kennedy, G.G. and Smitley, D.R. (1985). Dispersal. *In* Spider mites. their biology, natural enemies and control. Vol. 1A, W. Helle and M. W. Sabelis, *Eds*. Elsevier Science Publishers B.V., Amsterdam, The Netherlands**:** 233–242.

Kenyon, F., Welsh, M.*, et al.* (2003). Expressed sequence tag survey of gene expression in the scab mite *Psoroptes ovis* – allergens, proteases and free-radical scavengers. Parasitology 126: 451-460

Kerr, I.D., Haider, A.J.*, et al.* (2011). The ABCG family of membrane-associated transporters: you don't have to be big to be mighty. British Journal of Pharmacology 164: 1767-1779.

Khajehali, J., Van Nieuwenhuyse, P.*, et al.* (2011). Acaricide resistance and resistance mechanisms in *Tetranychus urticae* populations from rose greenhouses in the Netherlands. Pest Management Science 67: 1424-1433.

Khila, A. and Grbic, M. (2007). Gene silencing in the spider mite *Tetranychus urticae*: dsRNA and siRNA parental silencing of the Distal-less gene. Development Genes and Evolution 217: 241-251.

King-Jones, K., Horner, M.A.*, et al.* (2006). The DHR96 nuclear receptor regulates xenobiotic responses in *Drosophila*. Cell Metabolism 4: 37-48.

Kirkness, E.F., Haas, B.J.*, et al.* (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proceedings of the National Academy of Sciences of the United States of America 107: 12168-12173.

Kiss, T., Cadar, D.*, et al.* (2012a). Tick prevention at a crossroad: New and renewed solutions. Veterinary Parasitology 187: 357-366.

Kiss, K., Brozik, A.*, et al.* (2012b). Shifting the paradigm: the putative mitochondrial protein ABCB6 resides in the lysosomes of cells and in the plasma membrane of erythrocytes. PLoS One 7: e37378.

Kjer, K.M. (2004). Aligned 18S and insect phylogeny. Systematic Biology 53: 506-514.

Klimov, P.B. and Oconnor, B.M. (2008). Origin and higher-level relationships of psoroptidian mites (Acari : Astigmata : Psoroptidia): Evidence from three nuclear genes. Molecular Phylogenetics and Evolution 47: 1135-1156.

Klimov, P.B. and Oconnor, B.M. (2009). Improved tRNA prediction in the American house dust mite reveals widespread occurrence of extremely short minimal tRNAs in acariform mites. BMC Genomics 10: 598.

Klimov, P.B. and Knowles, L.L. (2011). Repeated parallel evolution of minimal rRNAs revealed from detailed comparative analysis. Journal of Heredity 102: 283-293.

Klimov, P.B. and OConnor, B. (2013). Is permanent parasitism reversible? – Critical evidence from early evolution of house dust mites. Systematic Biology: in press, doi: 10.1093/sysbio/syt1008.

Klompen, H., Lekvelshvili, M.*, et al.* (2007). Phylogeny of parasitiform mites (Acari) based on rRNA. Molecular Phylogenetics and Evolution 43: 936-951.

Kluge, A.G. (1989). A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes). Systematic Zoology 38: 7-25.

Knee, W., Beaulieu, F.*, et al.* (2012). Species boundaries and host range of tortoise Mites (Uropodoidea) phoretic on bark Beetles (Scolytinae), using morphometric and molecular markers. PLoS One 7: e47243.

Knipple, D.C. and Soderlund, D.M. (2010). The ligand-gated chloride channel gene family of *Drosophila melanogaster*. Pesticide Biochemistry and Physiology 97: 140-148.

Komoto, N., Quan, G.-X.*, et al.* (2009). A single-base deletion in an ABC transporter gene causes white eyes, white eggs, and translucent larval skin in the silkworm w-3(oe) mutant. Insect Biochemistry and Molecular Biology 39: 152-156.

Kovalchuk, A. and Driessen, A.J.M. (2010). Phylogenetic analysis of fungal ABC transporters. BMC Genomics 11: 177.

Krantz, G.W. and Lindquist, E.E. (1979). Evolution of phytophagous mites (Acari). Annual Review of Entomology 24: 121-158.

Krantz, G.W. and Walter, D.E. (2009). A manual of Acarology. Tech University Press, Texas: 704 p.

Kretschmer, M., Leroch, M., *et al.* (2009). Fungicide-driven evolution and molecular basis of multidrug resistance in field populations of the grey mould fungus *Botrytis cinerea*. PLoS Pathogens 5: e1000696.

Krieger, R.I., Feeny, P.P., *et al.* (1971). Detoxication enzymes in the guts of caterpillars: an evolutionary answer to plant defenses ? Science 172: 579-581.

Krieger, M., Smith, L.C., *et al.* (1979). Reconstituted low-density lipoprotein - vehicle for the delivery of hydrophobic fluorescent-probes to cells. Journal of Supramolecular Structure 10: 467-478.

Kruh, G.D. and Belinsky, M.G. (2003). The MRP family of drug efflux pumps. Oncogene 22: 7537-7552.

Kucharski, R., Maleszka, J., *et al.* (2008). Nutritional control of reproductive status in honeybees via DNA methylation. Science 319: 1827-1830.

Kumazawa, Y. and Endo, H. (2004). Mitochondrial genome of the Komodo dragon: Efficient sequencing method with reptile-oriented primers and novel gene rearrangements. DNA Research 11: 115-125.

Kuo, C.C., Huang, C.L., *et al.* (2011). Identification of potential hosts and vectors of scrub typhus and tick-borne spotted fever group rickettsiae in eastern Taiwan. Medical and Veterinary Entomology 25: 169-177.

Kurscheid, S., Lew-Tabor, A.E., *et al.* (2009). Evidence of a tick RNAi pathway by comparative genomics and reverse genetics screen of targets with known loss-of-function phenotypes in *Drosophila*. BMC Molecular Biology 10: 26.

Kuwana, H., Shimizu-Nishikawa, K., *et al.* (1996). Molecular cloning and characterization of the ABC transporter expressed in Trachea (ATET) gene from *Drosophila melanogaster*. Biochimica Et Biophysica Acta-Gene Structure and Expression 1309: 47-52.

Kwon, D.H., Yoon, K.S., *et al.* (2010). A point mutation in a glutamate-gated chloride channel confers abamectin resistance in the two-spotted spider mite, *Tetranychus urticae* Koch. Insect Molecular Biology 19: 583-591.

Kwon, D.H., Park, J.H., *et al.* (2013). Screening of lethal genes for feeding RNAi by leaf disc-mediated systematic delivery of dsRNA in *Tetranychus urticae*. Pesticide Biochemistry and Physiology 105: in press, doi: 10.1016/j.pestbp.2012.1012.1001.

Labbé, R., Caveney, S., *et al.* (2011). Genetic analysis of the xenobiotic resistance-associated ABC gene subfamilies of the Lepidoptera. Insect Molecular Biology 20: 243-256.

Laird, P.W. (2010). Principles and challenges of genome-wide DNA methylation analysis. Nature Reviews Genetics 11: 191-203.

Lamping, E., Baret, P.V., *et al.* (2010). Fungal PDR transporters: Phylogeny, topology, motifs and function. Fungal Genetics and Biology 47: 127-142.

Langmead, B., Trapnell, C., *et al.* (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10.

Lanning, C.L., Fine, R.L., *et al.* (1996). Tobacco budworm P-glycoprotein: biochemical characterization and its involvement in pesticide resistance. Biochimica Et Biophysica Acta-General Subjects 1291: 155-162.

Laslett, D. and Canback, B. (2008). ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24: 172-175.

Laux, V., Perzborn, E., *et al.* (2009). Direct inhibitors of coagulation proteins - the end of the heparin and low-molecular-weight heparin era for anticoagulant therapy? Thrombosis and Haemostasis 102: 892-899.

Lavrov, D.V., Brown, W.M., *et al.* (2000a). A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forficatus*. Proceedings of the National Academy of Sciences of the United States of America 97: 13738-13742.

Lavrov, D.V., Boore, J.L., *et al.* (2000b). The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. Molecular Biology and Evolution 17: 813-824.

Le Goff, G., Hamon, A., *et al.* (2005). Resistance to fipronil in *Drosophila simulans*: influence of two point mutations in the RDL GABA receptor subunit. Journal of Neurochemistry 92: 1295-1305.

Lee, Y.S., Oh, J., *et al.* (2008). Mitome: dynamic and interactive database for comparative mitochondrial genomics in metazoan animals. Nucleic Acids Research 36: D938-D942.

Lees, K. and Bowman, A.S. (2007). Tick neurobiology: recent advances and the post-genomic era. Invertebrate Neuroscience 7: 183-198.

Lespine, A., Ménez, C., *et al.* (2012). P-glycoproteins and other multidrug resistance transporters in the pharmacology of anthelmintics: Prospects for reversing transport-dependent anthelmintic resistance. International Journal for Parasitology: Drugs and Drug Resistance 2: 58-75.

Levinson, G. and Gutman, G.A. (1987). Slipped-strand misparing - A major mechanism for DNA-sequence evolution. Molecular Biology and Evolution 4: 203-221.

Lewis, D.L., Farr, C.L., *et al.* (1995). Drosophila melanogaster mitochondrial DNA: Completion of the nucleotide sequence and evolutionary comparisons. Insect Molecular Biology 4: 263-278.

Li, X., Zangerl, A., *et al.* (2000). Cross-resistance to a-cypermethrin after xanthotoxin ingestion in *Helicoverpa zea* (Lepidoptera: Noctuidae). Journal of Economic Entomology 93: 18-25.

Li, L., Stoeckert, C.J., *et al.* (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Research 13: 2178-2189.

Li, X., Baudry, J., *et al.* (2004). Structural and functional divergence of insect CYP6B proteins: From specialist to generalist cytochrome P450. Proceedings of the National Academy of Sciences of the United States of America 101: 2939-2944.

Li, A., Yang, Y., *et al.* (2006). Investigation of resistance mechanisms to fipronil in diamondback moth (Lepidoptera : Plutellidae). Journal of Economic Entomology 99: 914-919.

Li, X., Schuler, M., *et al.* (2007). Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. Annual Review of Entomology 52: 231-253.

Li, H., Liu, H., *et al.* (2012). The complete mitochondrial genome and novel gene arrangement of the unique-headed bug *Stenopirates* sp. (Hemiptera: Enicocephalidae). PLoS One 7: e29419.

Lilley, C.J., Davies, L.J., *et al.* (2012). RNA interference in plant parasitic nematodes: a summary of the current status. Parasitology 139: 630-640.

Lindquist, E.E. (1984). Current theories on the evolution of major groups of Acari and on their relationships with other groups of Arachnida with consequent implications for their classification. *In* Acarology, Vol 1. D.A. Griffiths and C.E. Bowman, *Eds.* Ellis Horwood Ltd., Chichester, UK: 28-62.

Lindquist, E.E. (1996). Phylogenetic relationships. *In* Eriophyoid mites - their biology, natural enemies and control., E. E. Lindquist, M. W. Sabelis and J. Bruin, *Eds*. Elsevier Science, Amsterdam**:** 301-327.

Lindquist, E.E., Bruin, J., *et al.* (1996). Eriophyoid mites: their biology, natural enemies and control. Elsevier Science, Amsterdam, The Netherlands**:** 822 p.

Lindquist, E.E. (2009). Classification. *In* A manual of Acarology. G.W. Krantz and D.E. Walter, *Eds.* Tech University Press, Texas, USA: 97-98.

Lindroth, R.L. (1989). Chemical ecology of the luna moth effects of host plant on detoxification enzyme-activity. Journal of Chemical Ecology 15: 2019-2029.

Liu, Z.W., Williamson, M.S., *et al.* (2005). A nicotinic acetylcholine receptor mutation conferring target-site resistance to imidacloprid in *Nilaparvata lugens* (brown planthopper). Proceedings of the National Academy of Sciences of the United States of America 102: 8420-8425.

Liu, B., Jiang, G., *et al.* (2011a). Analysis of transcriptome differences between resistant and susceptible strains of the citrus red mite *Panonychus citri* (Acari: Tetranychidae). PLoS One 6: e28516.

Liu, S., Zhou, S., *et al.* (2011b). Genome-wide identification and characterization of ATP-binding cassette transporters in the silkworm, *Bombyx mori*. BMC Genomics 12: 491.

Ljungren, E.L., Nilsson, D., *et al.* (2003). Expressed sequence tag analysis of *Sarcoptes scabiei*. Parasitology 127: 139-145.

Lowe, T.M. and Eddy, S.R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research 25: 955-964.

Lunt, D.H. and Hyman, B.C. (1997). Animal mitochondrial DNA recombination. Nature 387: 247-247.

Lyko, F. and Maleszka, R. (2011). Insects as innovative models for functional studies of DNA methylation. Trends in Genetics 27: 127-131.

Lynagh, T. and Lynch, J.W. (2010). A glycine residue essential for high ivermectin sensitivity in Cys-loop ion channel receptors. International Journal for Parasitology 40: 1477-1481.

Mackenzie, S.M., Brooker, M.R., *et al.* (1999). Mutations in the white gene of *Drosophila melanogaster* affecting ABC transporters that determine eye colouration. Biochimica Et Biophysica Acta-Biomembranes 1419: 173-185.

Magalhães, S., Fayard, J., *et al.* (2007). Adaptation in a spider mite population after long-term evolution on a single host plant. Journal of Evolutionary Biology 20: 2016-2027.

Magalhães, S., Blanchet, E., *et al.* (2009). Are adaptation costs necessary to build up a local adaptation pattern? BMC Evolutionary Biology 9: 182.

Magalhães, S., Blanchet, E., *et al.* (2011). Environmental effects on the detection of adaptation. Journal of Evolutionary Biology 24: 2653-2662.

Malone, J.H. and Oliver, B. (2011). Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biology 9: 34.

Mans, B.J., de Klerk, D., *et al.* (2012). The mitochondrial genomes of *Nuttalliella namaqua* (Ixodoidea: Nuttalliellidae) and *Argas africolumbae* (Ixodoidae: Argasidae): estimation of divergence dates for the major tick lineages and reconstruction of ancestral blood-feeding characters. PLoS One 7: e49461.

Mao, Y.-B., Tao, X.-Y., *et al.* (2011). Cotton plants expressing CYP6AE14 double-stranded RNA show enhanced resistance to bollworms. Transgenic Research 20: 665-673.

Marcadé, I., Cordaux, R., *et al.* (2007). Structure and evolution of the atypical mitochondrial genome of *Armadillidium vulgare* (Isopoda, Crustacea). Journal of Molecular Evolution 65: 651-659.

Marchler-Bauer, A., Lu, S., *et al.* (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Research 39: D225-D229.

Maritz-Olivier, C., Stutzer, C., *et al.* (2007). Tick anti-hernostatics: targets for future vaccines and therapeutics. Trends in Parasitology 23: 397-407.

Maserti, B.E., Del Carratore, R., *et al.* (2011). Comparative analysis of proteome changes induced by the two spotted spider mite *Tetranychus urticae* and methyl jasmonate in citrus leaves. Journal of Plant Physiology 168: 392-402.

Masta, S.E. (2000). Mitochondrial sequence evolution in spiders: Intraspecific variation in tRNAs lacking the T psi C arm. Molecular Biology and Evolution 17: 1091-1100.

Masta, S.E. and Boore, J.L. (2004). The complete mitochondrial genome sequence of the spider *Habronattus oregonensis* reveals rearranged and extremely truncated tRNAs. Molecular Biology and Evolution 21: 893-902.

Masta, S.E. and Boore, J.L. (2008). Parallel evolution of truncated transfer RNA genes in arachnid mitochondrial genomes. Molecular Biology and Evolution 25: 949-959.

Masta, S.E. (2010). Mitochondrial rRNA secondary structures and genome arrangements distinguish chelicerates: Comparisons with a harvestman (Arachnida: Opiliones: Phalangium opilio). Gene 449: 9-21.

Matera, I., Ferraroni, M., *et al.* (2010). Catechol 1,2-dioxygenase from the Gram-positive *Rhodococcus opacus* 1CP: Quantitative structure/activity relationship and the crystal structures of native enzyme and catechols adducts. Journal of Structural Biology 170: 548-564.

Matsuda, T., Hinomoto, N., *et al.* (2012). Molecular-based identification and phylogeny of *Oligonychus* species (Acari: Tetranychidae). Journal of Economic Entomology 105: 1043-1050.

May, P., Woldt, E., *et al.* (2007). The LDL receptor-related protein (LRP) family: An old family of proteins with new physiological functions. Annals of Medicine 39: 219-228.

Mayer, F. and Kerth, G. (2005). Microsatellite evolution in the mitochondrial genome of Bechstein's bat (*Myotis bechsteinii*). Journal of Molecular Evolution 61: 408-416.

Mayer, F., Mayer, N., *et al.* (2009). Evolutionary conservation of vertebrate blood-brain barrier chemoprotective mechanisms in *Drosophila*. Journal of Neuroscience 29: 3538-3550.

McCall, J., Alva, R., *et al.* (2004). Comparative efficacy of a combination of fipronil/(S)-methoprene, a combination of imidacloprid/permethrin, and imidacloprid against fleas and ticks when administered topically to dogs. International Journal of Applied Research in Veterinary Medicine 1: 74.

McDougall, P. (2008). Phillips McDougall Agriservice Report. Pathead, Midlothian.

McKenzie, J.A. and Batterham, P. (1994). The genetic, molecular and phenotypic consequences of selection for insecticide resistance. Trends in Ecology & Evolution 9: 166-169.

McQuilton, P., St Pierre, S.E., *et al.* (2012). FlyBase 101-the basics of navigating FlyBase. Nucleic Acids Research 40: D706-D714.

Meister, G. and Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. Nature 431: 343-349.

Merino, O., Almazan, C., *et al.* (2011). Control of *Rhipicephalus* (*Boophilus*) *microplus* infestations by the combination of subolesin vaccination and tick autocidal control after subolesin gene knockdown in ticks fed on cattle. Vaccine 29: 2248-2254.

Metzker, M.L. (2010). Applications of next generation sequencing: Sequencing technologies - the next generation. Nature Reviews Genetics 11: 31-46.

Meusemann, K., von Reumont, B.M., *et al.* (2010). A phylogenomic approach to resolve the arthropod tree of life. Molecular Biology and Evolution 27: 2451-2464.

Meyer, F., Flötenmeyer, M., *et al.* (2012). The sulfonylurea receptor SUR is dispensable for chitin synthesis in *Drosophila melanogaster* embryos. Pest Management Science, doi: 10.1002/ps.3476.

Michelmore, R.W., Paran, I., *et al.* (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis - a rapid method to detect markers in specific genomic regions by using segregating populations. Proceedings of the National Academy of Sciences of the United States of America 88: 9828-9832.

Migeon, A. and Dorkeld, F. (**2010**). Spider Mites Web: a comprehensive database for the Tetranychidae." from http://www.montpellier.inra.fr/CBGP/spmweb.

Mills, L.R. (1974). Structure of visual system of 2-spotted spider-mite, *Tetranychus urticae*. Journal of Insect Physiology 20: 795-808.

Misra, J.R., Horner, M.A., *et al.* (2011). Transcriptional regulation of xenobiotic detoxification in *Drosophila*. Genes & Development 25: 1796-1806.

Mita, K., Kasahara, M., *et al.* (2004). The genome sequence of silkworm, *Bombyx mori*. DNA Research 11: 27-35.

Mitani, H., Talbert, A., *et al.* (2004). New World relapsing fever *Borrelia* found in *Ornithodoros porcinus* ticks in central Tanzania. Microbiology and Immunology 48: 501-505.

Miyamoto, J., Ishii, A., *et al.* (1975). Successful method for mass-culture of house dust mite, *Dermatophagoides pteronyssinus* (Trouessart, 1897). Japanese Journal of Experimental Medicine 45: 133-138.

Monnier, V., Girardot, F., *et al.* (2002). Modulation of oxidative stress resistance in *Drosophila melanogaster* by gene overexpression**.** Genesis 34: 76-79.

Moran, N.A. and Jarvik, T. (2010). Lateral transfer of genes from fungi underlies carotenoid production in aphids. Science 328: 624-627.

Morita, M. and Imanaka, T. (2012). Peroxisomal ABC transporters: Structure, function and role in disease. Biochimica Et Biophysica Acta-Molecular Basis of Disease 1822: 1387-1396.

Morozova, O. and Marra, M.A. (2008). Applications of next-generation sequencing technologies in functional genomics. Genomics 92: 255-264.

Mounsey, K.E., Holt, D.C., *et al.* (2006). Identification of ABC transporters in *Sarcoptes scabiei*. Parasitology 132: 883-892.

Mounsey, K.E., Dent, J.A., *et al.* (2007). Molecular characterisation of a pH-gated chloride channel from *Sarcoptes scabiei*. Invertebrate Neuroscience 7: 149-156.

Mounsey, K.E., Willis, C., *et al.* (2012). Quantitative PCR-based genome size estimation of the astigmatid mites *Sarcoptes scabiei*, *Psoroptes ovis* and *Dermatophagoides pteronyssinus*. Parasites & Vectors 5: 3.

Mounsey, K.E., McCarthy, J.S., *et al.* (2013). Scratching the itch: new tools to advance understanding of scabies. Trends in Parasitology 29: 35-42.

Mul, M., van Niekerk, T., *et al.* (2009). Control methods for *Dermanyssus gallinae* in systems for laying hens: results of an international seminar. Worlds Poultry Science Journal 65: 589-599.

Mullin, C. and Croft, B. (1983). Host-related alterations of detoxification enzymes in *Tetranychus urticae* (Acari: Tetranychidae). Environmental Entomology 12: 1278-1282.

Mummery-Widmer, J.L., Yamazaki, M., *et al.* (2009). Genome-wide analysis of Notch signalling in *Drosophila* by transgenic RNAi. Nature 458: 987-U959.

Murray, C.L., Quaglia, M., *et al.* (1994). A putative nicotine pump at the metabolic blood-brain-barrier of the tobacco hornworm. Journal of Neurobiology 25: 23-34.

Murrell, A., Campell, N.J.H., *et al.* (2001a). A total-evidence phylogeny of ticks provides insights into the evolution of life cycles and biogeography. Molecular Phylogenetics and Evolution 21: 244-258.

Murrell, A., Campbell, N.J.H., *et al.* (2001b). Recurrent gains and losses of large (84-109 bp) repeats in the rDNA internal transcribed spacer 2 (ITS2) of rhipicephaline ticks. Insect Molecular Biology 10: 587-596.

Murrell, A., Dobson, S.J., *et al.* (2005). Relationships among the three major lineages of the Acari (Arthropoda : Arachnida) inferred from small subunit rRNA: paraphyly of the parasitiformes with respect to the opilioacariformes and relative rates of nucleotide substitution. Invertebrate Systematics 19: 383-389.

Nakao, T., Naoi, A., *et al.* (2010). Mutation of the GABA receptor associated with fipronil resistance in the whitebacked planthopper, *Sogatella furcifera*. Pesticide Biochemistry and Physiology 97: 262-266.

Narahashi, T., Zhao, X., *et al.* (2010). Glutamate-activated chloride channels: Unique fipronil targets present in insects but not in mammals. Pesticide Biochemistry and Physiology 97: 149-152.

Nardi, F., Carapelli, A., *et al.* (2001). The complete mitochondrial DNA sequence of the basal Hexapod *Tetrodontophora bielanensis*: Evidence for heteroplasmy and tRNA translocations. Molecular Biology and Evolution 18: 1293-1304.

Nava, S., Mastropaolo, M., *et al.* (2012). Mitochondrial DNA analysis of *Rhipicephalus sanguineus* sensu lato (Acari: Ixodidae) in the Southern Cone of South America. Veterinary Parasitology 190: 547-555.

Navajas, M., Lagnel, J., *et al.* (1998). Species-wide homogeneity of nuclear ribosomal ITS2 sequences in the spider mite *Tetranychus urticae* contrasts with extensive mitochondrial COI polymorphism. Heredity 80: 742-752.

Navajas, M., Le Conte, Y., *et al.* (2002). The complete sequence of the mitochondrial genome of the honeybee ectoparasite mite *Varroa destructor* (Acari : Mesostigmata). Molecular Biology and Evolution 19: 2313-2317.

Neiswander, C.R., Rodriguez, J.G., *et al.* (1950). Natural and induced variations in two-spotted spider mite populations. Journal of Economic Entomology 43: 633-636.

Nene, V. (2009). Tick genomics - coming of age. Frontiers in Bioscience 14: 2666-2673.

Nicastro, R., Sato, M., *et al.* (2010). Milbemectin resistance in *Tetranychus urticae* (Acari: Tetranychidae): selection, stability and cross-resistance to abamectin. Experimental and Applied Acarology 50: 231-241.

Nicastro, R.L., Sato, M.E., *et al.* (2011). Fitness costs associated with milbemectin resistance in the two-spotted spider mite *Tetranychus urticae*. International Journal of Pest Management 57: 223-228.

Nicholson, W.L., Sonenshine, D., *et al.* (2009). Ticks (Ixodidae). *In* Medical and veterinary entomology, G. Mullen and L. Durden, *Eds*. Academic Press, California, USA**:** 483-532.

Niehuis, O., Hartig, G., *et al.* (2012). Genomic and morphological evidence converge to resolve the enigma of Strepsiptera. Current Biology 22: 1309-1313.

Niu, J.Z., Dou, W., *et al.* (2012). Transcriptome analysis of the citrus red mite, *Panonychus citri,* and its gene expression by exposure to insecticide/acaricide. Insect Molecular Biology 21: 422-436.

Njue, A.I., Hayashi, J., *et al.* (2004). Mutations in the extracellular domains of glutamate-gated chloride channel alpha 3 and beta subunits from ivermectin-resistant *Cooperia oncophora* affect agonist sensitivity. Journal of Neurochemistry 89: 1137-1147.

Nygaard, S., Zhang, G., *et al.* (2011). The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. Genome Research 21: 1339-1348.

OConnor, B.M. (1984). Phylogenetic relationships among higher taxa in the Acariformes, with particular reference to the Astigmata. *In* Acarology VI, D. A. Griffiths and C. E. Bowman, *Eds*. Ellis-Horwood Ltd., Chichester**:** 19-27.

Oakeshott, J.G., Johnson, R.M., *et al.* (2010). Metabolic enzymes associated with xenobiotic and chemosensory responses in *Nasonia vitripennis*. Insect Molecular Biology 19: 147-163.

Oakley, T.H. (2003). On homology of arthropod compound eyes. Integrative and Comparative Biology 43: 522-530.

Ohtsuki, T., Watanabe, Y., *et al.* (2001). An "elongated" translation elongation factor Tu for truncated tRNAs in nematode mitochondria. Journal of Biological Chemistry 276: 21571-21577.

Ohtsuki, T. and Watanabe, Y. (2007). T-armless tRNAs and elongated elongation factor Tu. IUBMB Life 59: 68-75.

Ojala, D., Montoya, J., *et al.* (1981). Transfer-RNA punctuation model of RNA processing in human mitochondria. Nature 290: 470-474.

Ortells, M.O. and Lunt, G.G. (1995). Evolutionary history of the ligand-gated ion-channel superfamily of receptors. Trends in Neurosciences 18: 121-127.

Ovchinnikov, S. and Masta, S. (2012). Pseudoscorpion mitochondria show rearranged genes and genome-wide reductions of RNA gene sizes and inferred structures, yet typical nucleotide composition bias. BMC Evolutionary Biology 12: 31.

Ozawa, R., Nishimura, O., *et al.* (2012). Temperature-dependent, behavioural, and transcriptional variability of a tritrophic interaction consisting of bean, herbivorous mite, and predator. Molecular Ecology 21: 5624-5635.

Özvegy, C., Litman, T., *et al.* (2001). Functional characterization of the human multidrug transporter, ABCG2, expressed in insect cells. Biochemical and Biophysical Research Communications 285: 111-117.

Pachl, P., Domes, K., *et al.* (2012). Convergent evolution of defense mechanisms in oribatid mites (Acari, Oribatida) shows no "ghosts of predation past". Molecular Phylogenetics and Evolution 65: 412-420.

Palli, S.R., Bai, H., *et al.* (2012). Insect Genomics. *In* Insect molecular biology and biochemistry, L. I. Gilbert, *Ed*. Elsevier/Academic Press, Amsterdam**:** 1-29.

Pamilo, P., Viljakainen, L., *et al.* (2007). Exceptionally high density of NUMTs in the honeybee genome. Molecular Biology and Evolution 24: 1340-1346.

Pao, S.S., Paulsen, I.T., *et al.* (1998). Major facilitator superfamily. Microbiology and Molecular Biology Reviews 62: 1-34.

Paper, J.M., Scott-Craig, J.S., *et al.* (2007). Comparative proteomics of extracellular proteins in vitro and in planta from the pathogenic fungus *Fusarium graminearum*. Proteomics 7: 3171-3183.

Papetti, C., Lio, P., *et al.* (2007). Antarctic fish mitochondrial genomes lack ND6 gene. Journal of Molecular Evolution 65: 519-528.

Pascussi, J.-M., Gerbal-Chaloin, S., *et al.* (2008). The tangle of nuclear receptors that controls xenobiotic metabolism and transport: Crosstalk and consequences. Annual Review of Pharmacology and Toxicology 48: 1-32.

Patramool, S., Choumet, V., *et al.* (2012). Update on the proteomics of major arthropod vectors of human and animal pathogens. Proteomics 12: 3510-3523.

Pauchet, Y., Muck, A., *et al.* (2008). Mapping the larval midgut lumen proteorne of *Helicoverpa armigera*, a generalist herbivorous insect. Journal of Proteome Research 7: 1629-1639.

Paytubi, S., Wang, X., *et al.* (2009). ABC50 promotes translation initiation in mammalian cells. Journal of Biological Chemistry 284: 24061-24073.

Peelman, F., Labeur, C., *et al.* (2003). Characterization of the ABCA transporter subfamily: Identification of prokaryotic and eukaryotic members, phylogeny and topology. Journal of Molecular Biology 325: 259-274.

Pepato, A.R., da Rocha, C.E.F., *et al.* (2010). Phylogenetic position of the acariform mites: sensitivity to homology assessment under total evidence. BMC Evolutionary Biology 10: 235.

Perna, N.T. and Kocher, T.D. (1995). Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. Journal of Molecular Evolution 41: 353-358.

Peter, R.J., Van den Bossche, P., *et al.* (2005). Tick, fly, and mosquito control - Lessons from the past, solutions for the future. Veterinary Parasitology 132: 205-215.

Petersen, T.N., Brunak, S., *et al.* (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods 8: 785-786.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Research 29: e45.

Pierleoni, A., Martelli, P.L., *et al.* (2008). PredGPI: a GPI-anchor predictor. BMC Bioinformatics 9: 392.

Pimentel, D. (2009). Environmental and economic costs of the application of pesticides primarily in the United States. *In* Integrated pest management. Volume 1: Innovation-development process, R. Peshin and A. K. Dhawan, *Eds*. Springer**:** 89-111.

Pitino, M., Coleman, A.D., *et al.* (2011). Silencing of aphid genes by dsRNA feeding from plants. PLoS One 6: e25709.

Pohl, P.C., Klafke, G.M., *et al.* (2011). ABC transporter efflux pumps: A defense mechanism against ivermectin in *Rhipicephalus* (*Boophilus*) *microplus*. International Journal for Parasitology 41: 1323-1333.

Popovic, M., Zaja, R., *et al.* (2010). A novel ABC transporter: the first insight into zebrafish (*Danio rerio*) ABCH1. Marine Environmental Research 69: S11-S13.

Posada, D. and Crandall, K.A. (1998). MODELTEST: testing the model of DNA substitution. Bioinformatics 14: 817-818.

Potter, C. (1952). An improved laboratory apparatus for applying direct sprays and surface films, with data on the electrostatic charge on atomized spray fluids. Annals of Applied Biology 38: 1-12.

Price, D.R.G. and Gatehouse, J.A. (2008). RNAi-mediated crop protection against insects. Trends in Biotechnology 26: 393-400.

Qin, W., Neal, S.J., *et al.* (2005). Cold hardening and transcriptional change in *Drosophila melanogaster*. Insect Molecular Biology 14: 607-613.

Radolf, J.D., Caimano, M.J., *et al.* (2012). Of ticks, mice and men: understanding the dual-host lifestyle of Lyme disease spirochaetes. Nature Reviews Microbiology 10: 87-99.

Rambaldi, D. and Ciccarelli, F.D. (2009). FancyGene: dynamic visualization of gene structures and protein domain architectures on genomic loci. Bioinformatics 25: 2281-2282.

Ramsey, J.S., Rider, D.S., *et al.* (2010). Comparative analysis of detoxification enzymes in *Acyrthosiphon pisum* and *Myzus persicae*. Insect Molecular Biology 19: 155-164.

Raspotnig, G., Norton, R.A., *et al.* (2011). Oribatid mites and skin alkaloids in poison frogs. Biology Letters 7: 555-556.

Raz, T., Kapranov, P., *et al.* (2011). Protocol dependence of sequencing-based gene expression measurements. PLoS One 6: e19287.

Rector, A., R. Tachezy, *et al.* (2004). A sequence-independent strategy for detection and cloning of circular DNA virus genomes by using multiply primed rolling-circle amplification. Journal of Virology 78: 4993-4998.

Reddy, B.P.N., Prasad, G.B.K.S., *et al.* (2011). In silico analysis of glutathione S-transferase supergene family revealed hitherto unreported insect specific delta- and epsilon-GSTs and mammalian specific mu-GSTs in *Ixodes scapularis* (Acari: Ixodidae). Computational Biology and Chemistry 35: 114-120.

Reddy, V.S., Shlykov, M.A., *et al.* (2012). The major facilitator superfamily (MFS) revisited. FEBS Journal 279: 2022-2035.

Rees, D.C., Johnson, E., *et al.* (2009). ABC transporters: the power to change. Nature Reviews Molecular Cell Biology 10: 218-227.

Regier, J.C., Shultz, J.W., *et al.* (2010). Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. Nature 463: 1079-U1098.

Ribeiro, J.M.C., Alarcon-Chaidez, F., *et al.* (2006). An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks. Insect Biochemistry and Molecular Biology 36: 111-129.

Ricardo, S. and Lehmann, R. (2009). An ABC transporter controls export of a *Drosophila* germ cell attractant. Science 323: 943-946.

Rice, P., Longden, I., *et al.* (2000). EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics 16: 276-277.

Richards, S., Liu, Y., *et al.* (2005). Comparative genome sequencing of *Drosophila pseudoobscura*: Chromosomal, gene, and cis-element evolution. Genome Research 15: 1-18.

Richards, T.A., Dacks, J.B., *et al.* (2006). Evolution of filamentous plant pathogens: Gene exchange across eukaryotic kingdoms. Current Biology 16: 1857-1864.

Richards, S., Gibbs, R.A., *et al.* (2010). Genome Sequence of the pea aphid *Acyrthosiphon pisum*. PLoS Biology 8: e1000313.

Richards, E.H., Jones, B., *et al.* (2011a). Salivary secretions from the honeybee mite, *Varroa destructor*: effects on insect haemocytes and preliminary biochemical characterization. Parasitology 138: 602-608.

Richards, T.A., Soanes, D.M., *et al.* (2011b). Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. Proceedings of the National Academy of Sciences of the United States of America 108: 15258-15263.

Riesgo, A., Andrade, S.C.S., *et al.* (2012). Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. Frontiers in Zoology 9: 33.

Robinson, G.E., Hackett, K.J., *et al.* (2011). Creating a buzz about insect genomes. Science 331: 1386-1386.

Robinson, S.W., Herzyk, P., *et al.* (2013). FlyAtlas: database of gene expression in the tissues of *Drosophila melanogaster*. Nucleic Acids Research 41: D744-D750.

Rodriguez-Valle, M., Lew-Tabor, A., *et al.* (2010). Comparative microarray analysis of *Rhipicephalus* (*Boophilus*) *microplus* expression profiles of larvae pre-attachment and feeding adult female stages on *Bos indicus* and *Bos taurus* cattle. BMC Genomics 11: 437.

Rombauts, S., Zhurov, V., *et al.* (2013). Mite genomics, a detoxification story. Plant & Animal Genomes XXI Conference, January 12-16th, San Diego, USA.

Roopesh, K., Guyot, S., *et al.* (2010). Biotransformation of procyanidins by a purified fungal dioxygenase: Identification and characterization of the products using mass spectrometry. Process Biochemistry 45: 904-913.

Roopesh, K., Abhilash, J., *et al.* (2012). Dioxygenase from *Aspergillus fumigatus* MC8: molecular modelling and in silico studies on enzyme-substrate interactions. Molecular Simulation 38: 144-151.

Rosenheim, J.A., Johnson, M.W., *et al.* (1996). Biochemical preadaptations, founder events, and the evolution of resistance in arthropods. Journal of Economic Entomology 89: 263-273.

Rota-Stabelli, O., Yang, Z., *et al.* (2009). MtZoa: A general mitochondrial amino acid substitutions model for animal evolutionary studies. Molecular Phylogenetics and Evolution 52: 268-272.

Rota-Stabelli, O., Kayal, E., *et al.* (2010). Ecdysozoan mitogenomics: evidence for a common origin of the legged Invertebrates, the Panarthropoda. Genome Biology and Evolution 2: 425-440.

Rota-Stabelli, O., Daley, Allison C., *et al.* (2013). Molecular timetrees teveal a Cambrian colonization of land and a new scenario for ecdysozoan evolution. Current Biology 23: 392-398.

Roth, C.W., Holm, I., *et al.* (2003). Identification of the *Anopheles gambiae* ATP-binding cassette transporter superfamily genes. Molecules and Cells 15: 150-158.

Roush, R.T. and McKenzie, J.A. (1987). Ecological genetics of insecticide and acaricide resistance. Annual Review of Entomology 32: 361-380.

Rozen, S. and Skaletsky, H.J. (2000). Primer3 on the WWW for general users and for biologist programmers. *In* Bioinformatics Methods and Protocols: Methods in Molecular Biology, S. Krawetz and S. Misener, *Eds*. Humana Press, Totowa, New Jersey, USA**:** 365-386.

Rubinoff, D. and Holland, B.S. (2005). Between two extremes: Mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Systematic Biology 54: 952-961.

Rufener, L., Keiser, J., *et al.* (2010). Phylogenomics of ligand-gated ion channels predicts monepantel effect. PLoS Pathogens 6: e1001091.

Russel, F.G.M., Koenderink, J.B., *et al.* (2008). Multidrug resistance protein 4 (MRP4/ABCC4): a versatile efflux transporter for drugs and signalling molecules. Trends in Pharmacological Sciences 29: 200-207.

Saccone, C., Pesole, G., *et al.* (1991). The main regulatory region of mammalian mitochondrial DNA - Structure function model and evolutionary pattern. Journal of Molecular Evolution 33: 83-91.

Sachrajda, I. and Ratajewski, M. (2011). Mithramycin A suppresses expression of the human melanoma-associated gene ABCB8. Molecular Genetics and Genomics 285: 57-65.

Saidijam, M., Benedetti, G., *et al.* (2006). Microbial drug efflux proteins of the major facilitator superfamily. Current Drug Targets 7: 793-811.

Saier, M.H., Jr., Yen, M.R., *et al.* (2009). The Transporter Classification Database: recent advances. Nucleic Acids Research 37: D274-D278.

Saito, S., Tamura, K.*, et al.* (2005). Replication origin of mitochondrial DNA in insects. Genetics 171: 1695-1705.

Saldivar, L., Guerrero, F.D.*, et al.* (2008). Microarray analysis of acaricide-inducible gene expression in the southern cattle tick, *Rhipicephalus* (*Boophilus*) *microplus*. Insect Molecular Biology 17: 597-606.

Sambrook, J. and Russel, D. (1987). Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, United states**:** 1659 p.

Sanchez, D., Ganfornina, M.D.*, et al.* (2003). Exon-intron structure and evolution of the lipocalin gene family. Molecular Biology and Evolution 20: 775-783.

Sanchez, D., Ganfornina, M.D.*, et al.* (2006). Lipocalin genes and their evolutionary history. *In* Lipocalins, B. Åkerström, N. Borregaard, D. R. Flower and J. P. Salier, *Eds*. Landes Bioscience, Austin, Texas, USA**:** 5-16.

Sanger, F., Nicklen, S.*, et al.* (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America 74: 5463-5467.

Sarmento, R.A., Lemos, F.*, et al.* (2011). A herbivore that manipulates plant defence. Ecology Letters 14: 229-236.

Sasabe, M., Wen, Z.M.*, et al.* (2004). Molecular analysis of CYP321A1, a novel cytochrome P450 involved in metabolism of plant allelochernicals (furanocoumarins) and insecticides (cypermethrin) in *Helicoverpa zea*. Gene 338: 163-175.

Sato, M.E., Silva, M.Z.d.*, et al.* (2005). Abamectin resistance in *Tetranychus urticae* Koch (Acari: Tetranychidae): selection, cross-resistance and stability of resistance. Neotropical Entomology 34: 991-998.

Sau, A., Tregno, F.P.*, et al.* (2010). Glutathione transferases and development of new principles to overcome drug resistance. Archives of Biochemistry and Biophysics 500: 116-122.

Schinkel, A.H. and Jonker, J.W. (2003). Mammalian drug efflux transporters of the ATP binding cassette (ABC) family: an overview. Advanced Drug Delivery Reviews 55: 3-29.

Schnizler, K., Saeger, B.*, et al.* (2005). A novel chloride channel in *Drosophila melanogaster* is inhibited by protons. Journal of Biological Chemistry 280: 16254-16262.

Scott, C., Pandey, G.*, et al.* (2008). The enzymatic basis for pesticide bioremediation. Indian Journal of Microbiology 48: 65-79.

Sen Zeng, R., Wen, Z.*, et al.* (2007). Allelochemical induction of cytochrome P450 monooxygenases and amelioration of xenobiotic toxicity in *Helicoverpa zea*. Journal of Chemical Ecology 33: 449-461.

Shah, S., Yarrow, C.*, et al.* (2012). Insecticide detoxification indicator strains as tools for enhancing chemical discovery screens. Pest Management Science 68: 38-48.

Shani, N., Sapag, A.*, et al.* (1996). Characterization and analysis of conserved motifs in a peroxisomal ATP-binding cassette transporter. Journal of Biological Chemistry 271: 8725-8730.

Shao, R.F, Aoki, Y.*, et al.* (2004). The mitochondrial genomes of soft ticks have an arrangement of genes that has remained unchanged for over 400 million years. Insect Molecular Biology 13: 219-224.

Shao, R.F., Mitani, H.*, et al.* (2005a). Novel mitochondrial gene content and gene arrangement indicate illegitimate inter-mtDNA recombination in the chigger mite, *Leptotrombidium pallidum*. Journal of Molecular Evolution 60: 764-773.

Shao, R.F., Barker, S.C.*, et al.* (2005b). Evolution of duplicate control regions in the mitochondrial genomes of Metazoa: A case study with Australasian *Ixodes* ticks. Molecular Biology and Evolution 22: 620-629.

Shao, R.F., Barker, S.C.*, et al.* (2006). Molecular mechanisms for the variation of mitochondrial gene content and gene arrangement among chigger mites of the genus *Leptotrombidium* (Acari : Acariformes). Journal of Molecular Evolution 63: 251-261.

Shao, R.F, Zhu, X.-Q.*, et al.* (2012). Evolution of extensively fragmented mitochondrial genomes in the lice of humans. Genome Biology and Evolution 4: 1088-1101.

Sharp, P.M., Cowe, E.*, et al.* (1988). Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens* - A review on the considerable within species diversity. Nucleic Acids Research 16: 8207-8211.

Shayeghi, M., Latunde-Dada, G.O., *et al.* (2005). Identification of an intestinal heme transporter. Cell 122: 789-801.

Shendure, J. and Ji, H. (2008). Next-generation DNA sequencing. Nature Biotechnology 26: 1135-1145.

Sheps, J.A., Ralph, S., *et al.* (2004). The ABC transporter gene family of *Caenorhabditis elegans* has implications for the evolutionary dynamics of multidrug resistance in eukaryotes. Genome Biology 5: R15.

Shimomura, M., Yokota, M., *et al.* (2004). Roles of loop C and the loop B-C interval of the nicotinic receptor at subunit in its selective interactions with imidacloprid in insects. Neuroscience Letters 363: 195-198.

Shimomura, M., Yokota, M., *et al.* (2006). Role in the selectivity of neonicotinoids of insect-specific basic residues in loop D of the nicotinic acetylcholine receptor agonist binding site. Molecular Pharmacology 70: 1255-1263.

Sijen, T., Fleenor, J., *et al.* (2001). On the role of RNA amplification in dsRNA-triggered gene silencing. Cell 107: 465-476.

Simon, C., Frati, F., *et al.* (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain-reaction primers. Annals of the Entomological Society of America 87: 651-701.

Singh, S.N., Kumari, B., *et al.* (2013). Bacterial degradation of pyrene in minimal salt medium mediated by catechol dioxygenases: Enzyme purification and molecular size determination. Bioresource Technology 133: 293-300.

Skerratt, L.F., Campbell, N.J.H., *et al.* (2002). The mitochondrial 12S gene is a suitable marker of populations of *Sarcoptes scabiei* from wombats, dogs and humans in Australia. Parasitology Research 88: 376-379.

Skoracka, A., Smith, L., *et al.* (2010). Host-plant specificity and specialization in eriophyoid mites and their importance for the use of eriophyoid mites as biocontrol agents of weeds. Experimental and Applied Acarology 51: 93-113.

Smith, C.D., Zimin, A., *et al.* (2011a). Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). Proceedings of the National Academy of Sciences of the United States of America 108: 5673-5678.

Smith, C.R., Smith, C.D., *et al.* (2011b). Draft genome of the red harvester ant *Pogonomyrmex barbatus*. Proceedings of the National Academy of Sciences of the United States of America 108: 5667-5672.

Smith, Z.D. and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nature Reviews Genetics 14: 204-220.

Snyder, M. and Glendinning, J. (1996). Causal connection between detoxification enzyme activity and consumption of a toxic plant compound. Journal of Comparative Physiology A - Sensory Neural and Behavioral Physiology 179: 255-261.

Song, F., You, Z., *et al.* (2009). Specific loops D, E and F of nicotinic acetylcholine receptor beta 1 subunit may confer imidacloprid selectivity between *Myzus persicae* and its predatory enemy *Pardosa pseudoannulata*. Insect Biochemistry and Molecular Biology 39: 833-841.

Song, L. and Wang, W. (2013). Genomes and evolutionary genomics of animals. Current Zoology 59: 87-98.

Sparagano, O., Pavlicevic, A., *et al.* (2009). Prevalence and key figures for the poultry red mite *Dermanyssus gallinae* infections in poultry farm systems. Experimental and Applied Acarology 48: 3-10.

Spieksma, F.T.M. (1997). Domestic mites from an acarologic perspectives. Allergy 52: 360-368.

Stach, T., Braband, A., *et al.* (2010). Erosion of phylogenetic signal in tunicate mitochondrial genomes on different levels of analysis. Molecular Phylogenetics and Evolution 55: 860-870.

Staton, J.L., Daehler, L.L., *et al.* (1997). Mitochondrial gene arrangement of the horseshoe crab *Limulus polyphemus* L: Conservation of major features among arthropod classes. Molecular Biology and Evolution 14: 867-874.

Stone, B.F. (1968). A formula for determining degree of dominance in cases of monofactorial inheritence of resistance to chemicals. Bulletin of the World Health Organization 38: 325-326.

Storms, J.J.H. (1971). Some physiological effects of spider mite infestation on bean plants. Netherlands Journal of Plant Pathology 77: 154-167.

Strong, D.R., Lawton, J.H., *et al.* (1984). Insects on plants: community patterns and mechanisms. Blackwell Scientific, Oxford**:** p.

Strycharz, J.P. (2010). Polygenic resistance in the highly DDT-resistant 91-R strain of *Drosophila melanogaster* involves decreased penetration, increased metabolism and direct excretion of DDT, University of Massachussets. Department of Veterinary and Animal Sciences. Master of Science**:** 83 p.

Stumpf, N. and Nauen, R. (2002). Biochemical markers linked to abamectin resistance in *Tetranychus urticae* (Acari: Tetranychidae). Pesticide Biochemistry and Physiology 72: 111-121.

Sturm, A., Cunningham, P., *et al.* (2009). The ABC transporter gene family of *Daphnia pulex*. BMC Genomics 10: 170.

Sturn, A., Quackenbush, J., *et al.* (2002). GENESIS: cluster analysis of microarray data. Bioinformatics 18: 207-208.

Suen, G., Teiling, C., *et al.* (2011). The genome sequence of the leaf-cutter ant *Atta cephalotes* reveals insights into its obligate symbiotic lifestyle. PLoS Genetics 7: e1002007.

Swofford, D.L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), version 4.0b10. Sunderland, Massachusetts, Sinauer Asociates.

Szeri, F., Ilias, A., *et al.* (2009). The high turnover *Drosophila* multidrug resistance-associated protein shares the biochemical features of its human orthologues. Biochimica Et Biophysica Acta-Biomembranes 1788: 402-409.

Taanman, J.W. (1999). The mitochondrial genome: structure, transcription, translation and replication. Biochimica et Biophysica Acta-Bioenergetics 1410: 103-123.

Talavera, G. and Vila, R. (2011). What is the phylogenetic signal limit from mitogenomes? The reconciliation between mitochondrial and nuclear data in the Insecta class phylogeny. BMC Evolutionary Biology 11: 315.

Tamura, K., Peterson, D., *et al.* (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28: 2731-2739.

Tarnay, J.N., Szeri, F., *et al.* (2004). The dMRP/CG6214 gene of *Drosophila* is evolutionarily and functionally related to the human multidrug resistance-associated protein family. Insect Molecular Biology 13: 539-548.

Tarr, P.T., Tarling, E.J., *et al.* (2009). Emerging new paradigms for ABCG transporters. Biochimica Et Biophysica Acta-Molecular and Cell Biology of Lipids 1791: 584-593.

Tatematsu, K.-i., Yamamoto, K., *et al.* (2011). Positional cloning of silkworm white egg 2 (w-2) locus shows functional conservation and diversification of ABC transporters for pigmentation in insects. Genes to Cells 16: 331-342.

Thao, M.L., Baumann, L., *et al.* (2004). Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). BMC Evolutionary Biology 4: 25.

Thompson, J.N. (1994). The coevolutionary process. The University of Chicago Press, Chicago**:** 383 p.

Thompson, J.D., Higgins, D.G., *et al.* (1994). CLUSTAL-W - Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22: 4673-4680.

Torricelli, G., Carapelli, A., *et al.* (2010). High divergence across the whole mitochondrial genome in the "pan-Antarctic" springtail *Friesea grisea*: Evidence for cryptic species? Gene 449: 30-40.

Trautwein, M.D., Wiegmann, B.M., *et al.* (2012). Advances in insect phylogeny at the dawn of the postgenomic era. Annual Review of Entomology 57: 449-468.

Tsagkarakou, A., Van Leeuwen, T., *et al.* (2009). Identification of pyrethroid resistance associated mutations in the para sodium channel of the two-spotted spider mite *Tetranychus urticae* (Acari: Tetranychidae). Insect Molecular Biology 18: 583-593.

Turbeville, J.M., Pfeifer, D.M., *et al.* (1991). The phylogenetic status of arthropods, as inferred from 18s ribosomal-rna sequences. Molecular Biology and Evolution 8: 669-686.

Ullmann, A.J., Piesman, J., *et al.* (2003). A preliminary linkage map of the hard tick, *Ixodes scapularis*. Insect Molecular Biology 12: 201-210.

Ullmann, A.J., Lima, C.M.R., *et al.* (2005). Genome size and organization in the blacklegged tick, *Ixodes scapularis* and the Southern cattle tick, *Boophilus microplus*. Insect Molecular Biology 14: 217-222.

Vaillancourt, F.H., Bolin, J.T., *et al.* (2006). The ins and outs of ring-cleaving dioxygenases. Critical Reviews in Biochemistry and Molecular Biology 41: 241-267.

van Bronswijk, J.E.M.A. (1981). House dust biology. NIB Publishers, The Netherlands**:** 314 p.

Van de Peer, Y., Caers, A., *et al.* (1998). Database on the structure of small ribosomal subunit RNA. Nucleic Acids Research 26: 179-182.

van der Hammen, L. (1989). An introduction to comparative Arachnology. SPB Academic Publishing bv, The Hague, The Netherlands: 576 p.

Van Dongen, S. (2000). Graph clustering by flow simulation. Utrecht, University of Utrecht. PhD**:** 169 p.

Van Leeuwen, T., Stillatus, V., *et al.* (2004). Genetic analysis and cross-resistance spectrum of a laboratory-selected chlorfenapyr resistant strain of two-spotted spider mite (Acari : Tetranychidae). Experimental & Applied Acarology 32: 249-261.

Van Leeuwen, T., Van Pottelberge, S., *et al.* (2005). Comparative acaricide susceptibility and detoxifying enzyme activities in field-collected resistant and susceptible strains of *Tetranychus urticae*. Pest Management Science 61: 499-507.

Van Leeuwen, T., Vanholme, B., *et al.* (2008). Mitochondrial heteroplasmy and the evolution of insecticide resistance: Non-Mendelian inheritance in action. Proceedings of the National Academy of Sciences of the United States of America 105: 5980-5985.

Van Leeuwen, T., Vontas, J., *et al.* (2010a). Acaricide resistance mechanisms in the two-spotted spider mite *Tetranychus urticae* and other important Acari: A review. Insect Biochemistry and Molecular Biology 40: 563-572.

Van Leeuwen, T., Witters, J., *et al.* (2010b). The control of eriophyoid mites: state of the art and future challenges. Experimental and Applied Acarology 51: 205-224.

Van Leeuwen, T., Demaeght, P., *et al.* (2012). Population bulk segregant mapping uncovers resistance mutations and the mode of action of a chitin synthesis inhibitor in arthropods. Proceedings of the National Academy of Sciences of the United States of America 109: 4407-4412.

Van Leeuwen, T., Dermauw, W., *et al.* (2013). Spider mite control and resistance management: does a genome help? Pest Management Science 69: 156-159.

van Lenteren, J.C., Roskam, M.M., *et al.* (1997). Commercial mass production and pricing of organisms for biological control of pests in Europe. Biological Control 10: 143-149.

van Lenteren, J.C. (2012). The state of commercial augmentative biological control: plenty of natural enemies, but a frustrating lack of uptake. Biocontrol 57: 1-20.

Van Pottelberge, S., Van Leeuwen, T., *et al.* (2009). Resistance mechanisms to mitochondrial electron transport inhibitors in a field-collected strain of *Tetranychus urticae* Koch (Acari: Tetranychidae). Bulletin of Entomological Research 99: 23-31.

Van Zee, J.P., Geraci, N.S., *et al.* (2007). Tick genomics: The *Ixodes* genome project and beyond. International Journal for Parasitology 37: 1297-1305.

Veerman, A. and Helle, W. (1978). Evidence for functional involvement of carotenoids in photoperiodic reaction of spider mites. Nature 275: 234-234.

Veerman, A. (1985). Diapause. *In* Spider mites: their biology, natural enemies and control. Vol. 1A, M. W. Sabelis and W. Helle, *Eds*. Elsevier Science Publishers B.V., Amsterdam, The Netherlands**:** 279-316.

Verrier, P.J., Bird, D., *et al.* (2008). Plant ABC proteins - a unified nomenclature and updated inventory. Trends in Plant Science 13: 151-159.

Vetting, M.W. and Ohlendorf, D.H. (2000). The 1.8 angstrom crystal structure of catechol 1,2-dioxygenase reveals a novel hydrophobic helical zipper as a subunit linker. Structure with Folding and Design 8: 429-440.

Via, S. (1990). Ecological genetics and host adaptation in herbivorous insects - the experimental study of evolution in natural and agricultural systems. Annual Review of Entomology 35: 421-446.

Voorhorst, R., Spieksma, F.T., *et al.* (1967). House-dust mite (*Dermatophagoides pteronyssinus*) and allergens it produces. Identity with house-dust allergen. Journal of Allergy 39: 325-339.

Wade, N.M., Tollenaere, A.*, et al.* (2009). Evolution of a novel carotenoid-binding protein responsible for crustacean shell color. Molecular Biology and Evolution 26: 1851-1864.

Walker, D.W., Muffat, J.*, et al.* (2006). Overexpression of a *Drosophila* homolog of apolipoprotein D leads to increased stress resistance and extended lifespan. Current Biology 16: 674-679.

Walter, D.E. and Proctor, H.C. (1999). Mites. Ecology, evolution and behaviour. CABI Publishing, Wallingford**:** 322 p.

Walton, S.F., Dougall, A.*, et al.* (2004). Genetic epidemiology of *Sarcoptes scabiei* (Acari : Sarcoptidae) in northern Australia. International Journal for Parasitology 34: 839-849.

Wang, Z., Gerstein, M.*, et al.* (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10: 57-63.

Waterhouse, R.M., Zdobnov, E.M.*, et al.* (2011). OrthoDB: the hierarchical catalog of eukaryotic orthologs in 2011. Nucleic Acids Research 39: D283-D288.

Wauthy, G., Leponce, M.*, et al.* (1998). The backward jump of a box moss mite. Proceedings of the Royal Society B-Biological Sciences Series B 265: 2235-2242.

Webster, L.M.I., Thomas, R.H.*, et al.* (2004). Molecular systematics of *Acarus siro* s. lat., a complex of stored food pests. Molecular Phylogenetics and Evolution 32: 817-822.

Wei, S.-J., Shi, M.*, et al.* (2010). New views on strand asymmetry in insect mitochondrial genomes. PLoS One 5: e12708.

Wei, D.-D., Shao, R.*, et al.* (2012). The multipartite mitochondrial genome of *Liposcelis bostrychophila*: insights into the evolution of mitochondrial genomes in bilateral animals. PLoS One 7: e33973.

Weinstock, G.M., Robinson, G.E.*, et al.* (2006). Insights into social insects from the genome of the honeybee Apis mellifera. Nature 443: 931-949.

Werren, J.H., Richards, S.*, et al.* (2010). Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. Science 327: 343-348.

Whalon, M.E., Mota-Sanchez, R.M.*, et al.* (2013). Artrhopods Resistant to Pesticides Database (ARPD)." Retrieved 15th February 2013, from http://www.pesticideresistance.org.

Wheat, C.W. and Wahlberg, N. (2013). Phylogenomic Insights into the Cambrian explosion, the colonization of land and the evolution of flight in Arthropoda. Systematic Biology 62: 93-109.

Whitman, D.W. and Agrawal, A.A. (2009). What is phenotypic plasticity and why is it important. *In* Phenotypic plasticity of insects, D. W. Whitman and T. N. Ananthakrishnan, *Eds*. Science Publishers Inc, Enfield, New Hampshire**:** 1-63.

Willadsen, P. (2004). Anti-tick vaccines. Parasitology 129: S367-S387.

Wilson, L.J., Bauer, L.R.*, et al.* (1998). Effect of early season insecticide use on predators and outbreaks of spider mites (Acari : Tetranychidae) in cotton. Bulletin of Entomological Research 88: 477-488.

Wolstenholme, D.R., Macfarlane, J.L.*, et al.* (1987). Bizarre transfer-RNAs inferred from DNA-sequences of mitochondrial genomes of nematode worms. Proceedings of the National Academy of Sciences of the United States of America 84: 1324-1328.

Wolstenholme, D.R. (1992). Animal mitochondrial DNA: structure and evolution. International Review of Cytology 141: 173-216.

Wolstenholme, A.J. (2010). Recent progress in understanding the interaction between avermectins and ligand-gated ion channels: putting the pests to sleep. Invertebrate Neuroscience 10: 5-10.

Wu, Y.C. and Horvitz, H.R. (1998). The *C. elegans* cell corpse engulfment gene ced-7 encodes a protein similar to ABC transporters. Cell 93: 951-960.

Wurm, Y., Wang, J.*, et al.* (2011). The genome of the fire ant *Solenopsis invicta*. Proceedings of the National Academy of Sciences of the United States of America 108: 5679-5684.

Xia, X. and Xie, Z. (2001). DAMBE: Software package for data analysis in molecular biology and evolution. Journal of Heredity 92: 371-373.

Xia, X.H., Xie, Z.*, et al.* (2003). An index of substitution saturation and its application. Molecular Phylogenetics and Evolution 26: 1-7.

Xia, Q.Y., Zhou, Z.Y.*, et al.* (2004). A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). Science 306: 1937-1940.

Xie, X., Cheng, T., *et al.* (2012). Genome-wide analysis of the ATP-binding cassette (ABC) transporter gene family in the silkworm, *Bombyx mori*. Molecular Biology Reports 39: 7281-7291.

Xiong, J., Feng, L., *et al.* (2010). Genome-wide identification and evolution of ATP-binding cassette transporters in the ciliate *Tetrahymena thermophila*: A case of functional divergence in a multigene family. BMC Evolutionary Biology 10: 330.

Yamazaki, N., Ueshima, R., *et al.* (1997). Evolution of pulmonate gastropod mitochondrial genomes: Comparisons of gene organizations of *Euhadra*, *Cepaea* and *Albinaria* and implications of unusual tRNA secondary structures. Genetics 145: 749-758.

Yang, X.M., Margolies, D.C., *et al.* (2001). Host plant-induced changes in detoxification enzymes and susceptibility to pesticides in the twospotted spider mite (Acari : Tetranychidae). Journal of Economic Entomology 94: 381-387.

Yang, M.Y., Bowmaker, M., *et al.* (2002a). Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication. Cell 111: 495-505.

Yang, X.M., Buschman, L.L., *et al.* (2002b). Susceptibility and detoxifying enzyme activity in two spider mite species (Acari : Tetranychidae) after selection with three insecticides. Journal of Economic Entomology 95: 399-406.

Yang, F., Jensen, J.D., *et al.* (2012). Secretomics identifies *Fusarium graminearum* proteins involved in the interaction with barley and wheat. Molecular Plant Pathology 13: 445-453.

Yano, S., Takabayashi, J., *et al.* (2001). Trade-offs in performance on different plants may not restrict the host plant range of the phytophagous mite, *Tetranychus urticae*. Experimental and Applied Acarology 25: 371-381.

Yao, X., Song, F., *et al.* (2008). Amino acids within loops D, E and F of insect nicotinic acetylcholine receptor beta subunits influence neonicotinoid selectivity. Insect Biochemistry and Molecular Biology 38: 834-840.

Yasukawa, T., Yang, M.Y., *et al.* (2005). A bidirectional origin of replication maps to the major noncoding region of human mitochondrial DNA. Molecular Cell 18: 651-662.

Yates, J.R., Ruse, C.I., *et al.* (2009). Proteomics by mass spectrometry: approaches, advances, and applications. Annual Review of Biomedical Engineering 11: 49-79.

Yepiskoposyan, H., Egli, D., *et al.* (2006). Transcriptome response to heavy metal stress in *Drosophila* reveals a new zinc transporter that confers resistance to zinc. Nucleic Acids Research 34: 4866-4877.

Yoder, A.D., Irwin, J.A., *et al.* (2001). Failure of the ILD to determine data combinability for slow loris phylogeny. Systematic Biology 50: 408-424.

Yokobori, S., Fukuda, N., *et al.* (2004). Long-term conservation of six duplicated structural genes in cephalopod mitochondrial genomes. Molecular Biology and Evolution 21: 2034-2046.

Yoon, K.S., Strycharz, J.P., *et al.* (2011). Brief exposures of human body lice to sublethal amounts of ivermectin over-transcribes detoxification genes involved in tolerance. Insect Molecular Biology 20: 687-699.

Yoshiura K., Kinoshita A., *et al*: (2006). A SNP in the ABCC11 gene is the determinant of human earwax type. *Nat Genet* , 38:324-330.

You, M., Yue, Z., *et al.* (2013). A heterozygous moth genome provides insights into herbivory and detoxification. Nature Genetics 45: 220-225.

Yu, S.J., Berry, R.E., *et al.* (1979). Host plant stimulation of detoxifying enzymes in a phytophagous insect. Pesticide Biochemistry and Physiology 12: 280-284.

Yu, L.-L., Cui, Y.-J., *et al.* (2010). The ionotropic gamma-aminobutyric acid receptor gene family of the silkworm, *Bombyx mori*. Genome 53: 688-697.

Yuan, M.-L., Wei, D.-D., *et al.* (2010). The complete mitochondrial genome of the citrus red mite *Panonychus citri* (Acari: Tetranychidae): high genome rearrangement and extremely truncated tRNAs. BMC Genomics 11: 597.

Yue, G.H., Liew, W.C., *et al.* (2006). The complete mitochondrial genome of a basal teleost, the Asian arowana (*Scleropages formosus*, Osteoglossidae). BMC Genomics 7: 242.

Zardoya, R. and Meyer, A. (1996). The complete nucleotide sequence of the mitochondrial genome of the lungfish (*Protopterus dolloi*) supports its phylogenetic position as a close relative of land vertebrates. Genetics 142: 1249-1263.

Zardoya, R. and Meyer, A. (1998). Cloning and characterization of a microsatellite in the mitochondrial control region of the African side-necked turtle, *Pelomedusa subrufa*. Gene 216: 149-153.

Zarubica, A., Trompier, D.*, et al.* (2007). ABCA1, from pathology to membrane function. Pfluegers Archiv European Journal of Physiology 453: 569-579.

Zhan, S., Merlin, C.*, et al.* (2011). The monarch butterfly genome yields insights into long-distance migration. Cell 147: 1171-1185.

Zhang, S.D. and Odenwald, W.F. (1995). Misexpression of the white (omega) gene triggers male-made courtship in *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America 92: 5525-5529.

Zhang, D.X. and Hewitt, G.M. (1997). Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. Biochemical Systematics and Ecology 25: 99-120.

Zhang, S., Feany, M.B.*, et al.* (2009). Inactivation of *Drosophila* Huntingtin affects long-term adult functioning and the pathogenesis of a Huntington's disease model. Disease Models & Mechanisms 2: 247-266.

Zhang, Z.-Q. (2011). Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness. Magnolia Press, Auckland, New Zealand**:** 237 p.

Zhao, Z.Y., Sheps, J.A.*, et al.* (2004). Expression analysis of ABC transporters reveals differential functions of tandemly duplicated genes in *Caenorhabditis elegans*. Journal of Molecular Biology 344: 409-417.

Zheng, Y.C., Hirschberg, B.*, et al.* (2002). Identification of two novel *Drosophila melanogaster* histamine-gated chloride channel subunits expressed in the eye. Journal of Biological Chemistry 277: 2000-2005.

Zhou, A.M., Hassel, B.A.*, et al.* (1993). Expression cloning of 2-5a-dependent RNAse - a uniquely regulated mediator of interferon action. Cell 72: 753-765.

Zong, J., Yao, X.*, et al.* (2009). Evolution of the RNA-dependent RNA polymerase (RdRP) genes: Duplications and possible losses before and after the divergence of major eukaryotic groups. Gene 447: 29-39.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research 31: 3406-3415.

Zutz, A., Gompf, S.*, et al.* (2009). Mitochondrial ABC proteins in health and disease. Biochimica Et Biophysica Acta-Bioenergetics 1787: 681-690.

Zwier, M.V., Verhulst, E.C.*, et al.* (2012). DNA methylation plays a crucial role during early *Nasonia* development. Insect Molecular Biology 21: 129-138.

Summary

This study aims at delivering a modest contribution to Acari (post-)genomics, a relatively new field in Acari research. This was done, firstly, by elucidating the mitochondrial (mt) genome sequences of two economically important mites and secondly, by exploiting the availability of the recently deciphered nuclear genome sequence of the two-spotted spider mite *Tetranychus urticae*. This genome sequence was the starting point for the design of a whole genome gene expression microarray which was used to study xenobiotic adaptation in spider mites (Chapter IV) but also led to the characterization of *T. urticae* gene families either known to be involved in detoxification (Chapter V) or as possible target-sites of pesticides (Chapter VI)

The **mt genome** of the **European house dust mite *Dermatophagoides pteronyssinus*** is a circular DNA molecule of 14,203bp. It contains the complete set of 37 genes usually present in metazoan mt genomes (**Chapter II**). Its mt gene order differs considerably from that of other Acari mt genomes. Compared to the mt genome of the horseshoe crab *Limulus polyphemus,* which is considered as the ancestral arthropod mt pattern, only 11 of the 38 gene boundaries are conserved. Comparing the *D. pteronyssinus* mt gene order with those of other Acari also revealed that this genome character seems less useful for deduction of phylogenetic relationships between Acari superorders. The majority strand of the *D. pteronyssinus* has a 72.6% AT-content and a GC-and AT-skew of 0.194 and -0.199. These skews are the reverse of those normally observed in typical animal mt genomes. Next, a microsatellite was detected in a large non-coding region (286bp), which probably functions as the control region. Furthermore, almost all tRNA genes lack a T-arm provoking the formation of canonical cloverleaf tRNA-structures, and both rRNA genes are considerably reduced in size. Finally, a maximum likelihood and Bayesian inference analysis clustered, concordant with traditional views of Acari phylogeny, *D. pteronyssinus* with *Steganacarus magnus*, forming a sistergroup of the Trombidiformes.

The **mt genome** sequence of a predatory mite, ***Phytoseiulus persimilis*** is described in **Chapter III**. The 16,199bp genome (79.8% AT) contains the standard set of 13 protein coding and 24 RNA genes. Similar to *D. pteronyssinus*, the mt gene order is extremely reshuffled and GC- and AT-skews (0.222 and -0.62, respectively) are reversed to those found in most animals. In contrast to the mt genome of the closely related *Metaseiulus occidentalis* - which was reported to be unusually large (24,960bp), to lack *nad6* and *nad3* protein coding genes and to contain 22 tRNAs without T-arms - the genome of *P. persimilis* has all features

of a standard metazoan mt genome. Consequently, we performed additional experiments on the *M. occidentalis* mt genome. Our preliminary restriction digests and Southern hybridization data revealed that its size is smaller than previously reported. Next, it was demonstrated that *nad3* is in fact not absent and that at least 15 of the 22 tRNAs in the *M. occidentalis* mt genome can be folded into canonical cloverleaf structures similar to their counterparts in *P. persimilis*. Finally, a maximum likelihood and Bayesian inference phylogenetic analysis clustered, concordant with traditional views of Acari phylogeny, *P. persimilis* with *M. occidentalis*, forming together with *V. destructor* a sistergroup of the Ixodida within the superorder of the Parasitiformes.

The T. urticae *nuclear genome*

T. urticae is an extreme polyphagous pest which is able to feed on more than 1,100 host plants and has an extraordinary ability to develop resistance. Recently, the complete nuclear genome of this species was sequenced (Grbić *et al*. 2011). A **whole genome gene expression microarray** was developed to follow changes in *T. urticae* gene expression after host transfer and in pesticide-susceptible versus pesticide-resistant strains (**Chapter IV**). When mites from a pesticide susceptible strain propagated on bean were transferred to a more challenging host (tomato), transcriptional responses increased over time with about 7.5 % of genes differentially expressed after five generations. The types of genes that responded to transfer offer immediate insights into the molecular basis of *T. urticae*'s broad plant host range. Marked changes in gene expression were observed in many members of several large gene families previously implicated in the detoxification of harmful compounds, such as the P450 monooxygenase genes. More surprisingly, we observed changes in the expression of many genes belonging to families not usually associated with toxin resistance. Among the genes with the most striking expression differences were those belonging to the lipocalin family and the major facilitator superfamily. Further, genes encoding intradiol ring-cleavage dioxygenases responded strongly to host transfer. A phylogenetic analysis revealed that these genes were recently acquired from either bacteria or fungi via horizontal gene transfer by herbivorous mites. Their proliferation in the *T. urticae* genome, and strong transcriptional response to host transfer, is consistent with selection for an expanded "toolkit" for the metabolism of foreign compounds by spider mites, even though the biochemical targets of these enzymes are currently unknown. Strikingly, transcriptional profiles of tomato adapted mites resembled those of multipesticide strains and adaptation of susceptible mites to tomato

decreased the susceptibility to unrelated pesticide classes. These findings suggest key roles for both an expanded environmental response gene repertoire and transcriptional regulation in the life history of a generalist herbivore. Furthermore, they support a model whereby selection for the ability to mount a broad response to the diverse chemical defenses of plants predisposes the evolution of pesticide resistance in generalists.

T. urticae gene families either known to be involved in detoxification of xenobiotics or as possible target-sites of pesticides were also characterized. The **ABC protein family** is a well-known family involved in **detoxification of xenobiotics**, with most members of this family either directly transporting toxicants out of the cell or after conjugation with glutathione. A total of 103 ABC genes were identified in the *T. urticae* genome (**Chapter V**). This is the highest number of ABC genes discovered in a metazoan species to date. Within the *T. urticae* ABC gene set, all members of the eight currently described subfamilies (A to H) were detected. A phylogenetic analysis revealed that the high number of ABC genes in *T. urticae* is due primarily to lineage-specific expansions of ABC genes within the ABCC, ABCG and ABCH subfamilies. In particular, the ABCC subfamily harbors the highest number of *T. urticae* ABC genes (39). In a comparative genomic analysis, we found clear orthologous relationships between a subset of *T. urticae* ABC proteins and ABC proteins in both vertebrates and invertebrates known to be involved in fundamental cellular processes. These included members of the ABCB-half transporters, and the ABCD, ABCE and ABCF families. Furthermore, one-to-one orthologues could be distinguished between *T. urticae* proteins and human ABCC10, ABCG5 and ABCG8, the *Drosophila melanogaster* sulfonylurea receptor and ecdysone-regulated transporter E23. Finally, expression profiling revealed that mainly ABC genes in the ABCC, ABCG and ABCH subfamilies were differentially expressed in multi-pesticide resistant mite strains and/or in mites transferred to challenging (toxic) host plants.

The **cys-loop ligand gated ion channel (cysLGIC) family** are well known as **targets of pesticides**. The cysLGIC super family of *T. urticae* represents the largest arthropod cysLGIC super family described to date and the first characterised one within the group of chelicerates (**Chapter VI**). Genome annotation, phylogenetic analysis and comparison of the cysLGIC subunits with their counterparts in insects revealed that the *T. urticae* genome encodes for a high number of glutamate- and histamine-gated chloride channel genes (GluCl and HisCl) compared to insects. Three orthologues of the insect γ-aminobutyric acid (GABA)-gated

chloride channel gene *Rdl* were detected. Other cysLGIC groups, such as the nAChR subunits, are more conserved and have clear insect orthologues. By analysing sequences of known insecticide targets, remarkable similarities between insensitive (resistant) insect cysLGIC forms and *T. urticae* orthologues were found, which might be involved in the low toxicity of certain insecticides (such as neonicotinoids and fipronil) against spider mites. Finally, on top of a previously documented mutation, it was shown that a novel G326E mutation in Tu_GluCl3 was strongly associated with high levels of abamectin resistance.

In conclusion, in this dissertation both fundamental as applied aspects of Acari genomics were explored and new insights in mitochondrial genomes, host plant adaptation and pesticide resistance of mites were revealed.

Samenvatting

Dit doctoraat levert een bijdrage tot de genomica van Acari, een opkomend onderzoeksgebied binnen de acarologie, *i.e.* de studie van mijten en teken (Hoofstuk I). Eerst werd het mitochondriale genoom van twee economisch belangrijke mijten, de Europese huisstofmijt *Dermatophagoides pteronyssinus* en de roofmijt *Phytoseiulus persimilisis*, gekarakteriseerd (Hoofdstuk II en III). In de daarop volgende hoofdstukken werd optimaal gebruik gemaakt van het recent ontcijferde genoom van de bonenspintmijt *Tetranychus urticae*. Op basis van dit genoom werd een genexpressie-microarray ontworpen en gebruikt voor het onderzoeken van adaptatie van spintmijten aan xenobiotische (lichaamsvreemde) stoffen (Hoofdstuk IV). Daarnaast werd zowel een *T. urticae* genfamilie die een rol speelt in detoxificatie van xenobiotische stoffen (Hoofdstuk V) als een genfamilie waarvan sommige leden coderen voor de inwerkingsplaats van pesticiden (Hoofdstuk VI) volledig gekarakteriseerd.

Het mitochondriaal genoom van de Europese huisstofmijt *Dermatophagoides pteronyssinus* is een ringvormige DNA molecule en bestaat uit 14 203 baseparen (Hoofdstuk II). Dit genoom bevat de volledige set van 37 genen, die normaal voorkomt in dierlijke mitochondriale genomen. De rangschikking van de mitochondriale genen van *D. pteronyssinus* verschilt sterk van die van andere mijten en van die van de Atlantische degenkrab *Limulus polyphemus*, waarvan het mitochondriaal genoom algemeen wordt beschouwd als het prototype van een mitochondriaal genoom van een geleedpotige. Uit de vergelijking van de rangschikking van de mitochondriale genen van *D. pteronyssinus* met deze van andere mijten bleek dat de rangschikking van de genen echter niet bruikbaar was voor het ontrafelen van fylogenetische verbanden tussen de verschillende superordes binnen de Acari.

De streng van het mitochondriaal genoom van *D. pteronyssinus* die de meerderheid van de 37 mitochondriale genen bevat, heeft een AT-gehalte van 72,6 % en een GC- en AT-asymmetrie van respectievelijk 0,194 en -0,199. Deze asymmetrie-waarden zijn tegenovergesteld aan die van de meeste dierlijke mitochondriale genomen. Verder werd er ook een microsatelliet geïdentificeerd in een grote niet-coderende regio van dit mitochondriaal genoom. Daarnaast bleek dat bijna alle tRNA genen een T-arm misten, zodat deze de typische klaverbladstructuur niet konden vormen, en dat beide rRNAs aanzienlijk kleiner waren dan normaal. Tenslotte groepeerde een fylogenetische analyse *D. pteronyssinus* met de mosmijt *Steganacarus magnus,* wat de traditionele inzichten over verwantschappen binnen de Acari bevestigde.

Het mitochondriaal genoom van de roofmijt *Phytoseiulus persimilis* wordt beschreven in hoofdstuk III. Dit ringvorming genoom bestaat uit 16 199 baseparen, heeft een AT-gehalte van 79.8% en bevat 13 proteïne coderende en 24 RNA genen. Net zoals bij *D. pteronyssinus* verschilt de rangschikking van mitochondriale genen sterk van die van *L. polyphemus*. Daarnaast zijn de GC- en AT-asymmetrie (respectievelijk 0.222 en -0.062) waarden van de streng die codeert voor de meerderheid van de 37 mitochondriale genen tegenovergesteld aan die van de meeste dierlijke mitochondriale genomen. Het reeds door andere auteurs gepubliceerde mitochondriaal genoom van de nauw verwante roofmijt *Metaseiulus occidentalis* vertoont enkele opvallende verschillen in vergelijking met dat van *P. persimilis*: het is uitzonderlijk groot (24 960 baseparen), de *nad6* en *nad3* genen ontbreken en het heeft 22 tRNA genen zonder T-arm. In een poging om deze verschillen te verklaren werden er extra experimenten uitgevoerd op het mitochondriaal genoom van *M. occidentalis*. Met behulp van technieken als *restriction digest* en *Southern hybridisation* werd aangetoond dat het mitochondriaal genoom van *M. occidentalis* kleiner was dan eerder gerapporteerd. Daarnaast werd bewezen dat *nad3* wel aanwezig was in het mitochondriaal genoom van *M. occidentalis* en dat tenminste 15 van de 22 mitochondriale tRNA genen een gelijkaardige secundaire structuur (*i.e.* met T-arm, klaverbladstructuur) hadden als deze van *P. persimilis*. Uiteindelijk werd ook nog een fylogenetische analyse op basis van mitochondriale proteïne coderende genen uitgevoerd. Zoals enigszins verwacht, groepeerde deze analyse *P. persimilis* samen met *M. occidentalis* en ondersteunde ze de monofylie van de Mesostigmata.

Het *T. urticae* nucleair genoom

De bonenspintmijt *T. urticae* kan meer dan 1100 waardplanten aantasten en ontwikkelt zeer snel resistentie tegen een breed gamma aan pesticiden. Omwille van die kenmerken is deze mijtensoort wereldwijd een economisch belangrijke gewasbeschadiger. Met behulp van de recent ontcijferde sequentie van het nucleair genoom van deze mijt (Grbić *et al.* 2011) kunnen de moleculaire mechanismen die aan de basis liggen van deze eigenschappen nu ook onderzocht worden. Een genexpressie microarray werd ontwikkeld om verschillen in genexpressie tussen gevoelige en multiresistente *T. urticae* stammen op te sporen. Met deze microarray werden eveneens genexpressie veranderingen gevolgd in spintmijten die werden overgebracht van hun vaste waardplant (boon) naar een nieuwe, minder geschikte (meer toxische) waardplant (tomaat) (Hoofdstuk IV). Het aantal genen met een significant verschillende expressie steeg naarmate spintmijten langer verbleven op tomaat. Bij

spintmijten die reeds gedurende vijf generaties op tomaat leefden was de expressie van 1206 *T. urticae* genen (ongeveer 7.5% van alle *T. urticae* genen) significant verschillend ten opzichte van die van spintmijten die zich op boon ontwikkelden. Er werden ook grote verschillen in expressie waargenomen bij leden van genfamilies waarvan geweten is dat ze een rol spelen bij de detoxificatie van xenobiotische stoffen, zoals de cytochroom P450 mono-oxygenasen. Daarnaast werden ook veranderingen waargenomen bij leden van genfamilies die nog nooit eerder met detoxificatie in verband werden gebracht, zoals genen die coderen voor lipocalinen, de *Major Facilitator* superfamilie (MFS) en intradiol ring splitsende dioxygenasen (ID-RSD). Lipocalinen kunnen binden met kleine hydrofobe moleculen en zouden mogelijks ook kunnen interageren met plantmetabolieten. MFS genen coderen voor transporteiwitten die toxische stoffen uit de cel zouden kunnen transporteren terwijl ID-RSDs in staat zijn om aromatische ringstructuren, die o.a. voorkomen in plantmetabolieten en pesticiden, af te breken. Een fylogenetische analyse van ID-RSDs wees uit dat *T. urticae* deze hoogst waarschijnlijk heeft verworven via een horizontale overdracht uit bacteriën of schimmels. Ook al is het specifieke substraat van deze *T. urticae* ID-RSDs nog niet gekend, toch kunnen we er van uit gaan dat de proliferatie van deze genen in het *T. urticae* genoom en hun sterk gewijzigde genexpressie na het veranderen van waardplant, er duidelijk op wijzen dat er binnen deze spintmijten een selectie gebeurd is naar een uitgebreide set van middelen om xenobiotische stoffen af te breken en/of te verwijderen. Net zoals tussen mijten op boon en mijten geadapteerd aan tomaat, werden bij vergelijking van multiresistente en gevoelige mijtenstammen ook veel genen gevonden met een significant verschillende genexpressie. Opvallend was dat veel van die gewijzigde genen dezelfde waren als diegene die bij mijten geadapteerd aan tomaat een verschil in genexpressie vertoonden (P450 mono-oxygenasen, lipocalinen, MFS en de ID-RSDs), en bovendien volgden ze dezelfde trend. Omwille van deze gelijkenissen werd er ook getest of mijten geadapteerd aan tomaat minder gevoelig waren voor pesticiden. Hieruit bleek dat adaptatie op tomaat de gevoeligheid van spintmijten voor drie van de vijf geteste acariciden significant deed verminderen. Deze resultaten suggereren ook dat polyfage – op veel waardplanten voorkomende - herbivoren sneller resistentie kunnen ontwikkelen op gewasbeschermingsmiddelen, omdat de waardplanten waar ze op voorkomen reeds selecteren voor een bepaald set aan detoxificatie-enzymen. De initiële toxiciteit van een gewasbeschermingsmiddel is dan verschillend per waardplant, en kan in sommige gevallen dicht aanleunen bij de gebruiksdosis van een pesticide, wat zou kunnen leiden tot een versnelde selectie naar resistentie.

Eén van de genfamilies die goed gekend is voor zijn rol in het ontgiften van xenobiotische stoffen (drugs, pesticiden) is de ABC proteïne familie. De meeste leden van deze familie coderen voor eiwitten die zorgen voor transport van toxische stoffen, ofwel direct uit de cel, ofwel na conjugatie met glutathion. In het genoom van de bonenspintmijt werden 103 ABC genen geïdentificeerd (Hoofdstuk V). Dit is tot op vandaag het hoogste aantal ABC genen dat ooit werd geïdentificeerd in een dier. Onder deze 103 ABC genen werden leden van alle tot nu toe gekende subfamilies, A tot H, geïdentificeerd. Een fylogenetische analyse wees uit dat het hoge aantal ABC genen in *T. urticae* voornamelijk te wijten is aan *T. urticae* specifieke expansies van ABC genen in de ABCC, ABCG en ABCH subfamilies, waarbij de *T. urticae* ABCC subfamilie het hoogste aantal (39) ABC genen bevat. Daarnaast werden er ook duidelijk *T. urticae* orthologen gevonden van ABC proteïnes die een rol spelen in fundamentele cellulaire processen in gewervelde en ongewervelde dieren zoals de ABCB-half en ABCD transporters en de ABCE en ABCF proteïnes. Verder werden er ook orthologe verbanden gevonden tussen *T. urticae* proteïnes en ABCC10, ABCG5 en ABCG8 van de mens en de sulfonylurea receptor en ecdyson gereguleerde transporter E23 van de fruitvlieg *Drosophila melanogaster*. Uit een expressieanalyse bleek tenslotte dat voornamelijk ABCC, ABCG en ABCH genen een significante wijziging in genexpressie vertoonden in multiresistente mijten en in mijten die werden overgeplaatst naar een minder geschikte, meer toxische waardplant.

De cys-lus ligand-afhankelijke ionkanalen (cysLAIK) familie bevat leden die goed gekend zijn als inwerkingsplaats van pesticiden. De cysLAIK familie van *T. urticae* is de grootste onder de geleedpotigen en is de eerste cysLAIK familie die werd beschreven voor een lid van de Chelicerata (Hoofdstuk VI). Genoomannotatie in combinatie met een fylogenetische analyse wees uit dat het *T. urticae* genoom, in vergelijking met dat van insecten, een hoger aantal genen bevat die coderen voor glutamaat- en histamine-afhankelijke chloridekanalen. Daarnaast werden ook drie orthologen van de GABA receptor in insecten gevonden terwijl van andere cysLAIK leden zoals de nicotine-acetylcholinereceptor subfamilie een gelijkaardig aantal als dat in insecten werd teruggevonden. Bij het analyseren van sequenties van bekende inwerkingsplaatsen van insecticiden, werden opvallende overeenkomsten gevonden tussen ongevoelige/resistente vormen van eiwitten uit de cysLAIK familie en hun orthologen in verschillende *T. urticae* stammen. Dit zou onder meer kunnen verklaren waarom bepaalde insecticiden zoals neonicotinoiden en fipronil een lage toxiciteit hebben

voor spintmijten. Bovenop een eerder beschreven mutatie werd er tenslotte aangetoond dat een nieuwe mutatie (G326E) sterk gerelateerd was aan een hoog niveau van abamectine-resistentie.

In dit doctoraat werden zowel fundamentele als toegepaste aspecten van de genomica van Acari behandeld. Hierdoor werden nieuwe inzichten verkregen in mitochondriale genomen, waardplantadaptatie en pesticideresistentie van mijten.

Curriculum Vitae

1 PERSONAL DETAILS

Name	Wannes Dermauw
Mobile	+32(0)485754077
Email	wannes.dermauw@ugent.be, dermauwwannes@gmail.com
Born	Leuven (Belgium) on July 2nd, 1982
Nationality	Belgian
Address	Filips Van Cleeflaan 119 9000 Gent Belgium

2 WORK EXPERIENCE

November 1st 2006 - present: Ghent University, Teaching Assistant/PhD student at the department of Crop Protection, Laboratory of Agrozoology, Faculty of Bioscience Engineering, Ghent University

3 EDUCATION

2006 - present: Ghent University, Belgium
PhD at the Faculty of Bioscience Engineering, Department of Crop Protection.
Title**:** "Acari genomics: from phylogenetics to agricultural applications"

2005 - 2006: Ghent University, Belgium
Master in Business Administration, graduated with distinction

2000 - 2005: Ghent University, Belgium
Masters in Bioscience Engineering, Agricultural Sciences, graduated with great distinction, Master thesis: "Alternative applications of spinosad".

1995 - 2000: Sint-Jozefsinstituut, Kortrijk, Belgium
Science and Math 6h, laureate

4 SCIENTIFIC OUTPUT

Publications in peer-reviewed international journals:

Dermauw W, Osborne EJ, Clark RM, Grbić M, Tirry L, Van Leeuwen T. A burst of ABC genes in the genome of the polyphagous spider mite *Tetranychus urticae*. *BMC Genomics*, in press.

Pavlidi N, **Dermauw W**, Rombauts S, Chrisargiris A, Van Leeuwen T, Vontas J (2013). Analysis of the olive fruit fly *Bactrocera oleae* transcriptome and phylogenetic classification of the major detoxification gene families. PLoS One, in press.

Demaeght P, **Dermauw W**, Tsakireli D, Khajehali J, Nauen R, Tirry L, Vontas J, Lümmen P, Van Leeuwen T (2013). Molecular analysis of spirodiclofen resistance in *Tetranychus urticae*: CYP392E10 metabolizes spirodiclofen, but not its corresponding enol. *Insect Biochemistry and Molecular Biology*, 43: 544-554

Dermauw W, Wybouw N, Rombauts S, Menten B, Vontas J, Grbić M, Clark RM, Feyereisen R, & Van Leeuwen T (2013) A link between host plant adaptation and pesticide resistance in the polyphagous spider mite *Tetranychus urticae*. *Proceedings of the National Academy of Sciences of the United States of America* 110 (2):E113-E122.

Wybouw N, Balabanidou V, Ballhorn DJ, **Dermauw W**, Grbić M, Vontas J, & Van Leeuwen T (2012) A horizontally transferred cyanase gene in the spider mite *Tetranychus urticae* is involved in cyanate metabolism and is differentially expressed upon host plant change. *Insect Biochemistry and Molecular Biology* 42(12):881-889.

Van Nieuwenhuyse P, Demaeght P, **Dermauw W**, Khalighi M, Stevens C, Vanholme B, Tirry L, Lümmen P, & Van Leeuwen T (2012) On the mode of action of bifenazate: New evidence for a mitochondrial target site. *Pesticide Biochemistry and Physiology* 104(2):88-95.

Van Leeuwen T, **Dermauw W**, Grbić M, Tirry L, & Feyereisen R (2012) Spider mite control and resistance management: does a genome help? *Pest Management Science* 65 (2), 156-159.

Van Leeuwen T, Demaeght P, Osborne EJ, **Dermauw W**, Gohlke S, Nauen R, Grbić M, Tirry L, Merzendorfer H, & Clark R (2012) Population bulk segregant mapping uncovers resistance mutations and the mode of action of a chitin synthesis inhibitor in arthropods. *Proceedings of the National Academy of Sciences of the United States of America* 105 (12): 4407-4412

Dermauw W, Ilias A, Riga M, Tsagkarakou A, Grbić M, Tirry L, Van Leeuwen T, & Vontas J (2012) The cys-loop ligand-gated ion channel gene family of *Tetranychus urticae*: implications for acaricide toxicology and a novel mutation associated with abamectin resistance. *Insect Biochemistry and Molecular Biology* 42(7):455-465.

Van Leeuwen T, Van Nieuwenhuyse P, Vanholme B, **Dermauw W**, Nauen R, & Tirry L (2011) Parallel evolution of cytochrome b mediated bifenazate resistance in the citrus red mite *Panonychus citri*. *Insect Molecular Biology* 20(1):135-140.

Grbić M, Van Leeuwen T, Clark RM, Rombauts S, Rouze P, Grbic V, Osborne EJ, **Dermauw W**, Phuong CTN, Ortego F*, et al.* (2011) The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* 479(7374):487-492.

Van Leeuwen T, Vontas J, Tsagkarakou A, **Dermauw W**, & Tirry L (2010) Acaricide resistance mechanisms in the two-spotted spider mite *Tetranychus urticae* and other important Acari: A review. *Insect Biochemistry and Molecular Biology* 40(8):563-572.

Dermauw W, Vanholme B, Tirry L, & Van Leeuwen T (2010) Mitochondrial genome analysis of the predatory mite *Phytoseiulus persimilis* and a revisit of the *Metaseiulus occidentalis* mitochondrial genome. *Genome* 53(4):285-301

Machtelinckx T, Van Leeuwen T, Vanholme B, Gehesquiere B, **Dermauw W**, Vandekerkhove B, Gheysen G, & De Clercq P (2009) Wolbachia induces strong cytoplasmic incompatibility in the predatory bug *Macrolophus pygmaeus*. *Insect Molecular Biology* 18(3):373-381.

Dermauw W, Van Leeuwen T, Vanholme B, & Tirry L (2009) The complete mitochondrial genome of the house dust mite *Dermatophagoides pteronyssinus* (Trouessart): a novel gene arrangement among arthropods. *BMC Genomics* 10.

Van Leeuwen T, Van de Veire M, **Dermauw W**, & Tirry L (2006) Systemic toxicity of spinosad to the greenhouse whitefly *Trialeurodes vaporariorum* and to the cotton leaf worm *Spodoptera littoralis*. *Phytoparasitica* 34(1):102-108.

Van Leeuwen T, **Dermauw W**, Van de Veire M, & Tirry L (2005) Systemic use of spinosad to control the two-spotted spider mite (Acari : Tetranychidae) on tomatoes grown in rockwool. *Experimental and Applied Acarology* 37(1-2):93-105.

Meetings with oral presentation

Dermauw W, Wybouw N, Rombauts S, Grbic M, Clark RM, Menten B, Feyereisen R, Tirry L & Van Leeuwen T (2012) Exploiting different host plants: a spider mite genomic and transcriptome perspective. XXIVth International congress of entomology, 19-24 August, Daegu, Korea

Dermauw W, Wybouw N, Vontas J, Grbic M, Rombauts S, Feyereisen R, Tirry L & Van Leeuwen T (2011) Whole genome gene expression microarray analysis of acaricide resistance in spider mites: involvement of new gene families. Resistance 2011, 5-7 September, Rothamsted Research, Harpenden, UK

Dermauw W, Vanholme B, Tirry L & Van Leeuwen T (2010) The mitochondrial genomes of *Phytoseiulus persimilis* and *Metaseiulus occidentalis*: more similar than different?, International Congress of Acarology, 23-27 August , Recife, PE, Brazil.

Van Leeuwen T, **Dermauw W**, Vanholme B, Gheysen G, Tirry L. (2008) Amplification of mite mitochondrial genomes, 6th Symposium of the European Association of Acarologists, 21-25 July, Montpellier, France.

Contributions to meetings with oral presentation:

Rombauts SM, Zhurov V, **Dermauw W**, Cao PT, Clark R, Van Leeuwen T, Van de Peer Y, Grbic M (2013) Mite Genomics, a Detoxification story. Plant & Animal Genomes XXI Conference, 12-16 January, San Diego, CA, USA.

Van Leeuwen T, **Dermauw W**, Wybouw N, Rombauts S, Menten B, Tirry L & Feyereisen R (2012) Genomic signatures for polyphagous feeding and pesticide resistance in Tetranychus urticae. 7th Symposium of the European Association of Acarologists, 9-13 July, Vienna, Austria.

Ilias A, **Dermauw W**, Riga M, Tsagkarakou A, Van Leeuwen T & Vontas J (2012) Characterization of abamectin resistance in *Tetranychus urticae* from Greece. 7th Symposium of the European Association of Acarologists, 9-13 July, Vienna, Austria.

Van Leeuwen T, Demaeght P, **Dermauw W**, Vontas J, Lümmen P, Feyereisen R & Tirry L (2012) P450s in the polyphagous chelicerate *Tetranychus urticae*: from genome annotation to functional expression. 11th Symposium on Cytochrome P450 Biodiversity and Biotechnology, 22-26 June, Torino, Italy.

Van Leeuwen T, **Dermauw W**, Rombauts S, Vontas J, Grbic M, Feyereisen R & Tirry L (2011) The xenobiotic metabolism in the spider mite *Tetranychus urticae*: whole genome microarray expression analysis identifies the involvement of new gene families. Sixth Symposium on Molecular Insect Science, 2-5 October, Amsterdam, The Netherlands.

Van Leeuwen T, **Dermauw W**, Van de Veire M & Tirry L (2005) Systemic use of spinosad to control spider mites, 57th International symposium on Crop protection, 10 May, Ghent Belgium.

Posters at national/international congresses:

Vontas, J, Chrisargiris A, Pavlidi N, Lyrakis E, **Dermauw W**, Van Leeuwen T (2012) Genomic approaches for understanding xenobiotic detoxification mechanisms of the olive fruit fly *Bactrocera oleae* and its adaptation to olives. 2nd International Symposium of TEAM, 3-6 July, Kolymbari, Crete, Greece.

Rombauts SM, Clark RM, Van Leeuwen T, Rouzé P, Cao PT, Osborne EJ, Bonnet E, Martens C, **Dermauw W**, Grbic V, Van de Peer Y, Grbic M (2011) The first complete Chelicerate Genome: *Tetranychus urticae*. Plant & Animal Genomes XIX Conference, 15-19 January, San Diego, CA, USA.

Vontas J, **Dermauw W**, Riga M, Ilias A, Tsagkarakou A, Van Leeuwen T. Molecular and functional characterization of insecticide/acaricide resistance in Tetranychus urticae from Greece (2011) Sixth Symposium on Molecular Insect Science, 2-5 October, Amsterdam, The Netherlands.

Publications in popular magazines:

Over polyfagie en resistentie (2013). Knack, 9 januari 2013, p.74.

Geen pantser meer (2012). Knack, 18 april 2012, p.83.

Etende spintmijt op een bonenblad (2012). Knack, 12 januari 2012, p. 77.

Balade aux pays des acariens. Mailleux A-C, Astudillo A, Clotuche G, Le Goff G, **Dermauw W** and Van Leeuwen T (2011). Parcs et réserves 66: 4-16.

Websites:

Mailleux AC, Clotuche G, **Dermauw W**, Astudillo AF, Goff G & Van Leeuwen T (2010) http://www.acari.be (website introducing mites and ticks to a broad audience)